Equilateral sets in finite-dimensional normed spaces

Konrad J. Swanepoel[*]

Abstract

This is an expository paper on the largest size of equilateral sets in finite-dimensional normed spaces.

Contents

1 Introduction

In this paper I discuss some of the more important known results on equilateral sets in finite-dimensional normed spaces with an analytical flavour. I have omitted a discussion of some of the low-dimensional results of a discrete geometric nature. The omission of a topic has no bearing on its importance, only on my prejudices and ignorance. The following topics are outside the scope of this paper: almost-equilateral sets, k-distance sets, and equilateral sets in infinite dimensional spaces. However, see the last section for references to these topics.

I have included the proofs that I find fascinating. The arguments are mostly very geometrical, using results from convex geometry and the local theory of Banach spaces, and also from algebraic topology. In the case of the ℓ_p-norm the known results use tools from linear algebra, probability theory, combinatorics and approximation theory.

Equilateral sets have been applied in differential geometry to find minimal surfaces in finite-dimensional normed spaces [24, 28].

Throughout the paper I state problems. Perhaps some of them are easy, but I do not know the answer to any of them.

1.1 Definitions

A subset S of a metric space (X, d) is λ-*equilateral* $(\lambda > 0)$ if $d(x, y) = \lambda$ for all $x, y \in S, x \neq y$. Let $e_\lambda(X)$ denote the largest size of a λ-equilateral set in X, if it exists, otherwise set $e_\lambda(X) = \infty$. If X is compact then it is easily seen that $e_\lambda(X) < \infty$ for each $\lambda > 0$.

In a normed space X, if S is λ-equilateral, then $\frac{\mu}{\lambda} S$ is μ-equilateral for any $\mu > 0$; thus the specific value of λ does not matter, and we only write $e(X)$. If X is finite-dimensional, then $e(X) < \infty$. In the remainder of the paper we abbreviate finite-dimensional normed space to *Minkowski space*. The *Banach-Mazur distance* between two Banach spaces X and Y is the infimum of all $c \geq 1$ such that there exists a linear isomorphism $T : X \to Y$ satisfying $\|T\|\|T^{-1}\| \leq c$. In the finite-dimensional case, this infimum is of course attained. We always use n to denote the dimension of X, and m the size of a set. We use $\mathbb{R}$ for the field of real numbers, $\mathbb{R}^n$ for the n-dimensional real vector space of n-dimensional column vectors, and $\boldsymbol{e}_i$ for the ith *standard basis vector* which has 1 in position i and 0 in all other positions. The transpose of a matrix A is denoted by A^{tr}. For a set $S \subseteq \mathbb{R}^n$, $\mathrm{conv}(S), \mathrm{int}(S), \mathrm{vol}(S)$ denotes its convex hull, interior, and Lebesgue measure (if S is measurable), respectively. By $\log x$ we denote the natural logarithm, and by $\lg x$ the logarithm with base 2.

1.2 Kusner's questions for the ℓ_p norms

On $\mathbb{R}^n$ the ℓ_p norm is defined by

$$\|(x_1, x_2, \ldots, x_n)\|_p := \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}$$

if $1 < p < \infty$, and for $p = \infty$ by

$$\|(x_1, x_2, \ldots, x_n)\|_\infty := \max\{|x_i| : i = 1, \ldots, n\}.$$

We denote the Minkowski space $(\mathbb{R}^n, \|\cdot\|_p)$ by ℓ_p^n. The ℓ_p distance between two points $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$ is $\|\boldsymbol{a} - \boldsymbol{b}\|_p$. Thus Euclidean space is ℓ_2^n.

The following lower bounds for $e(\ell_p^n)$ are simple:

- Since $\{\pm \boldsymbol{e}_i : i = 1, \ldots, n\}$ is 2-equilateral in ℓ_1^n, we have $e(\ell_1^n) \geq 2n$.

- For any $p \in (1, \infty)$ and an appropriate choice of $\lambda = \lambda(p) \in \mathbb{R}$, the set

$$\left\{ \boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n, \lambda \sum_{i=1}^n \boldsymbol{e}_i \right\}$$

is $2^{1/p}$-equilateral in ℓ_p^n; hence $e(\ell_p^n) \geq n + 1$.

- $\{(\varepsilon_1, \ldots, \varepsilon_n) : \varepsilon_i = \pm 1\}$ is 2-equilateral in ℓ_∞^n; thus $e(\ell_\infty^n) \geq 2^n$.

A theorem of Petty [29] (see Section 2.1) states that $e(X) \leq 2^n$ if the dimension of X is n; this then gives $e(\ell_\infty^n) = 2^n$. Kusner [20] asked whether $e(\ell_1^n) = 2n$ and $e(\ell_p^n) = n + 1$ holds for all $1 < p < \infty$. For p an even integer it is easy to show an upper bound of about pn using linear algebra (observation of Galvin; see Smyth [33]). It is possible to improve this upper bound to about $pn/2$, and in the particular case of $p = 4$ to show $e(\ell_4^n) = n + 1$ [37] (see Section 6.1). For $1 < p < 2$ there are examples known showing $e(\ell_p^n) > n + 1$ if n is sufficiently large (depending on p) [37] (see Section 3.2). On the other hand if p is sufficiently near 2 (depending on n), then $e(\ell_p^n) = n + 1$. Smyth [33] gives a quantitative estimate (see Section 4.2).

The above results are not difficult; to improve the 2^n upper bound for $e(\ell_p^n)$ turned out to need deeper results from analysis and linear algebra. The first breakthrough came when Smyth [33] combined a linear algebra method with the Jackson theorems from approximation theory to prove $e(\ell_p^n) < c_p n^{(p+1)/(p-1)}$ for $1 < p < \infty$. Then Alon and Pudlák [1] combined Smyth's method with a result on the rank of approximations of the identity matrix (the rank lemma; see Section 6.2) to prove $e(\ell_p^n) < c_p n^{(2p+2)/(2p-1)}$ for $1 \leq p < \infty$. For the ℓ_1 norm Alon and Pudlák combined the rank lemma with a probabilistic argument (randomized rounding) to prove $e(\ell_1^n) < cn \log n$, and more generally $e(\ell_p^n) < c_p n \log n$ if p is an odd integer.

Finally we mention that it is known that $e(\ell_1^3) = 6$ (Bandelt et al. [3]) and $e(\ell_1^4) = 8$ (Koolen et al. [23]). We do not discuss the proofs of these two results.

Problem 1 (Kusner). *Prove (or disprove) that $e(\ell_1^n) = 2n$ for $n \geq 5$.*

Problem 2 (Kusner). *Prove (or disprove) that $e(\ell_p^n) = n + 1$ for $p > 2$ ($p \neq 4$).*

Problem 3. *Prove (or disprove) that $e(\ell_p^n) < cn$ for $1 < p < 2$.*

2 General upper bounds

We now need the following notions from convex geometry. A *polytope* is a convex body that is the convex hull of finitely many points. A *face* of a convex body C is a set $F \subseteq C$ satisfying the following property:

If a segment $\boldsymbol{ab} \subseteq C$ intersects S, then $\boldsymbol{ab} \subseteq S$.

The following are well-known (and easily proved) facts from convex geometry:

1. Every point of a convex body C is contained in the relative interior of a unique face of C.

2. The faces of a polytope forms a ranked finite lattice, with dimension as rank function.

3. The 1-dimensional faces are called *vertices*, and the $(n-1)$-dimensional faces *facets*.

4. Any $(n-2)$-dimensional face is contained in exactly two facets.

A *parallelotope* is a translate of the polytope $\{\sum_{i=1}^n \lambda_i \boldsymbol{a}_i : -1 \leq \lambda \leq 1\}$ where $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ are linearly independent. It is easily seen that if a Minkowski space has a parallelotope as unit ball, then it is isometric to ℓ_∞^n.

2.1 The theorem of Petty and Soltan

The following theorem was proved by Petty [29], and later also by P. Soltan [34]. A different proof, using the isodiametric inequality, is given in [15].

3

Theorem 1 (Petty [29] & Soltan [34]). *For any n-dimensional normed space X, $e(X) \leq 2^n$ with equality iff X is isometric to ℓ_∞^n, in which case any equilateral set of size 2^n is the vertex set of some ball (which is then a parallelotope).*

Proof. Let $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_m \in X$ be 1-equilateral, where $m = e(X)$. Let $P = \operatorname{conv}\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_m\}$ be the convex hull of the points $\boldsymbol{p}_i$. If P is not full-dimensional we may use induction on n to obtain $m \leq 2^{n-1} < 2^n$ (and the cases $n = 0$ and $n = 1$ are trivial).

Thus without loss of generality P is full-dimensional. For each $i = 1, \ldots, m$, let $P_i = \frac{1}{2}(P + \boldsymbol{p}_i)$. Note the following easily proved facts:

1. Each $P_i \subseteq P$.

2. Each $P_i \subseteq B(\boldsymbol{p}_i, \frac{1}{2})$.

3. For any $i \neq j$, $P_i \cap P_j$ has empty interior.

Thus the P_i *pack* P, and we now calculate volumes:

$$\sum_{i=1}^m \operatorname{vol}(P_i) = \operatorname{vol}(\bigcup_{i=1}^m P_i) \leq \operatorname{vol}(P).$$

Since $\operatorname{vol}(P_i) = (\frac{1}{2})^n \operatorname{vol}(P)$, we obtain $m \leq 2^n$.

If equality holds, P is a polytope in $\mathbb{R}^n$ that is *perfectly packed* or *tiled* by 2^n translates of $\frac{1}{2}P$. By the following lemma P must be a parallelotope $\boldsymbol{v} + \{\sum_{i=1}^n \lambda_i \boldsymbol{a}_i : -1 \leq \lambda \leq 1\}$ with vertex set $\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_m\} = \boldsymbol{v} + \{\sum_{i=1}^n \varepsilon_i \boldsymbol{a}_i : \varepsilon_i = \pm 1\}$. Since $\boldsymbol{p}_i - \boldsymbol{p}_j$ is a unit vector for $i \neq j$, it follows that $\sum_{i=1}^n 2\varepsilon_i \boldsymbol{a}_i$ is a unit vector for all choices of $\varepsilon_i \in \{-1, 0, 1\}$ with not all $\varepsilon_i = 0$. It is then easy to see that the unit ball is the parallelotope $\{\sum_{i=1}^n \lambda_i \boldsymbol{a}_i : -1 \leq \lambda \leq 1\}$ (exercise), and the linear map sending $\boldsymbol{a}_i \mapsto \boldsymbol{e}_i$ is an isometry. $\quad\square$

Lemma 1 (Groemer [17]). *Let P be the convex hull of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{2^n} \in \mathbb{R}^n$. Suppose that $P = \bigcup_{i=1}^{2^n} \frac{1}{2}(P + \boldsymbol{v}_i)$. Then P is a parallelotope with vertices $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{2^n}$.*

Proof. Exercise. $\quad\square$

2.2 Strictly convex norms

A normed space is strictly convex if $\|\boldsymbol{x} + \boldsymbol{y}\| < \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ for all linearly independent $\boldsymbol{x}, \boldsymbol{y}$, or equivalently, if the boundary of the uniy ball does not contain a line segment. If X is a strictly convex Minkowski space, then $e(X) \leq 2^n - 1$ by Theorem 1, but as far as I know there is no better upper bound for $n \geq 4$.

Problem 4. *If X is an n-dimensional strictly convex Minkowski space, $n \geq 4$, prove that $e(X) \leq 2^n - 2$.*

In Section 3.1 we'll see that $e(X) = 3$ when $\dim X = 2$, and for $\dim X = 3$, Petty [29] deduced $e(X) \leq 5$ from a result of Grünbaum [19].

Conjecture 1 (Füredi, Lagarias, Morgan [15]). *There exists some constant $\varepsilon > 0$ such that if X is a strictly convex n-dimensional Minkowski space then $e(X) \leq (2 - \varepsilon)^n$.*

There exist strictly convex spaces with equilateral sets of size at least exponential in the dimension.

Theorem 2 (Füredi, Lagarias, Morgan [15]). *For each $n \geq 271$ there exists an n-dimensional strictly convex X with an equilateral set of size at least 1.058^n.*

In [15] the lower bound 1.02^n was obtained; however, their argument easily gives the bound as stated here. They constructed a norm $\|\cdot\|$ on $\mathbb{R}^n$ with the property that $\|\cdot\| - \varepsilon\|\cdot\|_2$ is still a norm, and asked whether the norm can in addition be C^∞ on $\mathbb{R}^n \setminus \{0\}$. In our construction the norm will lack the above

property, but it will be C^∞. However, most likely the norm in the above theorem can be made to have the above property and be C^∞.

We need the following very nice high-dimensional phenomenon. It is a special case of the Johnson-Lindenstrauss flattening lemma [22], although this special case is essentially the Gilbert-Varshamov lower bound for binary codes (see [25].)

Lemma 2. *For each $\delta > 0$ there exist $\varepsilon = \varepsilon(\delta) > 0$ and $n_0 = n_0(\delta) \geq 1$ such that for all $n \geq n_0$ there exist $m > (1 + \varepsilon)^n$ vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m \in \mathbb{R}^n$ satisfying*

$$\begin{cases} \langle \boldsymbol{w}_i, \boldsymbol{w}_i \rangle = 1 & \text{for all } i, \\ |\langle \boldsymbol{w}_i, \boldsymbol{w}_j \rangle| < \delta & \text{for all distinct } i, j. \end{cases} \tag{1}$$

We may take $\varepsilon = \delta^2/2$ and $n_0 \geq (120 \log 2)/(25\delta^4 - \delta^6)$.

We need the Chernoff inequality (proved in e.g. [25, Theorem 1.4.5]). The *binary entropy function* is defined by $H(0) = H(1) := 0$ and

$$H(x) := -x \lg x - (1 - x) \lg(1 - x), \text{ for } 0 < x < 1.$$

Lemma 3. *For any $\varepsilon > 0$ and $n \in \mathbb{N}$,*

$$\sum_{0 \leq k \leq \varepsilon n} \binom{n}{k} \leq 2^{nH(\varepsilon)}. \tag{2}$$

Proof of Lemma 2. For $n \geq 1$, define a graph on $\{-1, 1\}^n$ as vertex set by joining two points if the number of coordinates in which they differ is $\leq \frac{1}{2}(1 - \delta)n$ or $\geq \frac{1}{2}(1 + \delta)n$. This graph is regular of degree

$$m := \sum_{0 \leq k \leq \frac{1}{2}(1-\delta)n} \binom{n}{k} + \sum_{\frac{1}{2}(1+\delta)n \leq k \leq n} \binom{n}{k} \leq -1 + 2^{nH((1-\delta)/2)+1},$$

by (2). If $\boldsymbol{x}$ and $\boldsymbol{y}$ are not connected in this graph, then $n^{-1/2}\boldsymbol{x}$ and $n^{-1/2}\boldsymbol{y}$ are unit vectors satisfying $|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| < \delta$. We therefore have to show that the graph contains an *independent set* (a set without an edge between any two vertices) of at least $(1 + \epsilon)^n$ points. We arbitrarily choose a vertex, delete it and its m neighbours, choose a remaining vertex, delete it and its at most m neighbours, and continue until nothing remains. Thus we obtain an independent set of size at least

$$2^n/(m + 1) \leq 2^{n(1-H((1-\delta)/2))-1}$$

vertices, which is greater than $(1 + \delta^2/2)^n$ if

$$nf(\delta) > 1,$$

where $f(\delta) = 1 - H((1 - \delta)/2) - \lg(1 + \delta^2/2)$. But this is true for sufficiently large n, since it is easily seen that $f(\delta) > 0$ for $0 < \delta < 1$. That $n \geq (120 \log 2)/(25\delta^4 - \delta^6)$ is sufficient can be seen by calculating the Taylor expansion of $f(\delta)$. $\square$

Remarks

1. The proof shows that if $M(n, \delta)$ is the maximum number of unit vectors in $\mathbb{R}^n$ satisfying (1), then

$$\lim_{n \to \infty} \frac{1}{n} \lg M(n, \delta) \geq 1 - H\left(\frac{1 - \delta}{2}\right) = \frac{1}{\log 2}\left(\frac{\delta^2}{2} + \frac{\delta^4}{12} + \cdots\right).$$

If one repeatedly choose unit vectors on the Euclidean unit sphere $\mathbb{S}^{n-1}$ and delete spherical caps around them, then, as shown by Wyner [40, pp. 1089–1092] (slightly improving an earlier estimate of Shannon [32]),

$$\lim_{n \to \infty} \frac{1}{n} \lg M(n, \delta) \geq \lg(1 - \delta^2)^{-1/2} = \frac{1}{\log 2} \Big(\frac{\delta^2}{2} + \frac{\delta^4}{4} + \dots \Big),$$

which is the same up to first order, but has a better higher order term. Thus the above greedy argument is not really improved by choosing vectors from the Euclidean unit sphere.

2. Recently, the Gilbert-Varshamov lower bound for binary codes has been slightly improved [21], which will improve the lower bound in Lemma 2 to $c(\delta)n(1 + \delta^2/2)^n$.

3. If we prove this lemma by choosing vectors in $\{\pm 1\}$ randomly, and estimating the probabilities in the obvious way, then we obtain a slightly worse dependence of $\varepsilon = \delta^2/4$.

Proof of Theorem 2. By Lemma 2 there exists $\boldsymbol{w}_1, \dots, \boldsymbol{w}_m$, $m \geq 1.058^n$, satisfying (1) with $\delta = 1/3$.

We now show that $P = \text{conv}\{\boldsymbol{w}_i - \boldsymbol{w}_j : 1 \leq i, j \leq m\}$ is a centrally symmetric polytope with $m(m-1)$ vertices $\boldsymbol{w}_i - \boldsymbol{w}_j$, $i \neq j$. It is sufficient to note that the hyperplane

$$\{\boldsymbol{x} \in \mathbb{R}^n : \langle \boldsymbol{x}, \boldsymbol{w}_i - \boldsymbol{w}_j \rangle = \langle \boldsymbol{w}_i - \boldsymbol{w}_j, \boldsymbol{w}_i - \boldsymbol{w}_j \rangle\}$$

is a supporting hyperplane of P passing only through $\boldsymbol{w}_i - \boldsymbol{w}_j$. This follows from

$$\langle \boldsymbol{w}_j - \boldsymbol{w}_i, \boldsymbol{w}_i - \boldsymbol{w}_j \rangle, \pm \langle \boldsymbol{w}_i - \boldsymbol{w}_k, \boldsymbol{w}_i - \boldsymbol{w}_j \rangle,$$
$$\pm \langle \boldsymbol{w}_j - \boldsymbol{w}_k, \boldsymbol{w}_i - \boldsymbol{w}_j \rangle, \pm \langle \boldsymbol{w}_k - \boldsymbol{w}_\ell, \boldsymbol{w}_i - \boldsymbol{w}_j \rangle$$
$$\leq \quad \langle \boldsymbol{w}_i - \boldsymbol{w}_j, \boldsymbol{w}_i - \boldsymbol{w}_j \rangle \text{ for distinct } i, j, k, \ell,$$

which follow easily from (1) and $\delta = 1/3$.

With P as a unit ball we thus already have $\{\boldsymbol{w}_i : i = 1, \dots, m\}$ as an equilateral set. We now need to find a strictly convex centrally symmetric body with the vertices of P on its boundary. This is provided by the next lemma. $\qquad\square$

The following lemma is obvious, but we provide a simple proof.

Lemma 4. *Let S be the vertex set of a centrally symmetric polytope in $\mathbb{R}^n$. Then there is a smooth, strictly convex norm $\|\cdot\|$ on $\mathbb{R}^n$ such that $\|\boldsymbol{x}\| = 1$ for all $\boldsymbol{x} \in S$.*

Proof. Let $S = \{\pm\boldsymbol{x}_1, \dots, \pm\boldsymbol{x}_m\}$. For each $\boldsymbol{x}_i$, choose $\boldsymbol{y}_i \in \mathbb{R}^n$ such that $\langle \boldsymbol{y}_i, \boldsymbol{x}_i \rangle = 1$ and $|\langle \boldsymbol{y}_i, \boldsymbol{x}_j \rangle| < 1$ for all $i \neq j$. The required norm will be

$$\|\boldsymbol{x}\| = \Big(\sum_{j=1}^m \lambda_j |\langle \boldsymbol{y}_j, \boldsymbol{x} \rangle|^p \Big)^{1/p}$$

for suitably chosen $\lambda_j > 0$ and $1 < p < \infty$, i.e., we want to imbed S into the unit sphere of ℓ_p^m for some p.

Choose p large enough such that

$$|\langle \boldsymbol{y}_j, \boldsymbol{x}_i \rangle|^p < \frac{1}{2m} \text{ for all } j \neq i.$$

Consider the matrix $A = \big[a_{ij}\big]_{i,j=1}^m$ with

$$a_{ij} = \begin{cases} |\langle \boldsymbol{y}_j, \boldsymbol{x}_i \rangle|^p & \text{if } i \neq j, \\ 0 & \text{if } i = j. \end{cases}$$

Considering A as a linear transformation on ℓ_∞^m, we have $\|A\| < \frac{1}{2}$. We want to find a vector $\boldsymbol{v} = (\lambda_1, \ldots, \lambda_m)^{\mathrm{tr}} > 0$ such that

$$(I + A)\boldsymbol{v} = (1, \ldots, 1)^{\mathrm{tr}},$$

since then we would have $\|\boldsymbol{x}_i\| = 1$ for all i. However, since $\|A\| < \frac{1}{2}$, $I + A$ is invertible, and then necessarily $\boldsymbol{v} = (I + A)^{-1}(1, \ldots, 1)^{\mathrm{tr}}$. Also,

$$
\begin{aligned}
\|\boldsymbol{v} - (1, \ldots, 1)^{\mathrm{tr}}\|_\infty &= \|(I + A)^{-1}(1, \ldots, 1)^{\mathrm{tr}} - (1, \ldots, 1)^{\mathrm{tr}}\|_\infty \\
&\leq \|(I + A)^{-1} - I\| \|(1, \ldots, 1)\|_\infty \\
&= \|\sum_{i=1}^{\infty} A^i\| \leq \sum_{i=1}^{\infty} \|A\|^i < 1.
\end{aligned}
$$

Thus, $\lambda_i > 0$ for all i. $\qquad\square$

Note that in the above lemma we may choose p to be an even integer which gives a norm that is C^∞ on $\mathbb{R}^n \setminus \{\boldsymbol{o}\}$.

3 Simple lower bounds

We now consider the problem of finding equilateral sets in a general Minkowski space, thus providing a lower bound to $e(X)$.

3.1 General Minkowski spaces

Proposition 1. *If* $\dim X \geq 2$ *then* $e(X) \geq 3$.

Proof. Exercise. (Hint: Euclid Book I Proposition 1.) $\qquad\square$

The above proposition combined with Theorem 1 gives the following

Corollary 1. *If* $\dim X = 2$ *then*

$$
e(X) = \begin{cases} 4 & \text{if the unit ball of } X \text{ is a parallelogram,} \\ 3 & \text{if the unit ball of } X \text{ is not a parallelogram.} \end{cases}
$$

In general one would hope for the following conjecture, stated in [18, 28, 29, 39].

Conjecture 2. *If* $\dim X = n$ *then* $e(X) \geq n + 1$.

As seen above this is simple for $n = 2$, and the next theorem shows its truth for $n = 3$. However, for each $n \geq 4$ this is open, and the best that is known is the theorem of Brass and Dekster giving $e(X) \geq c(\log n)^{1/3}$, discussed in Section 5.

Theorem 3 (Petty [29]). *If* $\dim X \geq 3$ *then any equilateral set of size* 3 *may be extended to an equilateral set of size* 4.

The proof needs the following technical but simple lemma.

Lemma 5. *Let* S_2 *be the unit circle (i.e. boundary of the unit ball) of a* 2-*dimensional Minkowski space. Fix any* $\boldsymbol{u} \in S_2$. *Let* $\boldsymbol{f} : [0, 1] \to S_2$ *be a parametrization of an arc of* S_2 *from* $\boldsymbol{u}$ *to* $-\boldsymbol{u}$. *Fix* $\boldsymbol{p} = \lambda\boldsymbol{u}$, $\lambda > 0$. *Then* $t \mapsto \|\boldsymbol{p} - \boldsymbol{f}(t)\|$ *is strictly increasing before it reaches the value* $\|\boldsymbol{p} + \boldsymbol{u}\|$, *i.e. on the interval* $[0, t_0]$, *where* $t_0 = \min\{t : \|\boldsymbol{p} - \boldsymbol{f}(t)\| < \|\boldsymbol{p} + \boldsymbol{u}\|\}$.

Proof. Exercise. $\qquad\square$

Proof of Theorem 3. Let $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ be 1-equilateral in X. Without loss of generality $\boldsymbol{a} = \boldsymbol{o}$. Let $X_2 = \text{span}\{\boldsymbol{b}, \boldsymbol{c}\}$, and let X_3 be any 3-dimensional subspace of X containing X_2. Let S_2 be the unit circle of X_2, and let S_3 be the unit sphere of X_3.

Consider the mapping $\boldsymbol{f} : S_3 \to \mathbb{R}^2$ defined by $\boldsymbol{f}(x) = (\|\boldsymbol{x} - \boldsymbol{b}\|, \|\boldsymbol{x} - \boldsymbol{c}\|)$. We have to show that $(1,1)$ is in the range of $\boldsymbol{f}$. If we can show that the restriction of $\boldsymbol{f}$ to S_2 encircles $(1,1)$, then the theorem follows from the following topological argument.

S_2 can be contracted to a point on S_3, since S_3 is simply connected; thus $\boldsymbol{f}(S_2)$ can also be contracted to a point; if $(1,1)$ is in the interior of the Jordan curve $\boldsymbol{f}(S_2)$, then at some stage of the contraction the curve must pass through $(1,1)$, since $\mathbb{R}^2 \setminus \{(1,1)\}$ is not simply connected.

Thus we assume without loss of generality that $(1,1) \notin \boldsymbol{f}(S_2)$, and we now follow the curve $\boldsymbol{f}(S_2)$. We start at $\boldsymbol{f}(\boldsymbol{b}) = (0,1)$ and go to $\boldsymbol{f}(\boldsymbol{c}) = (1,0)$. By Lemma 5 the x-coordinate increases strictly while the y-coordinate decreases strictly. Therefore the arc of the curve strictly between $(0,1)$ and $(1,0)$ is contained in the square $0 < x, y < 1$. Then from $(1,0)$ to $\boldsymbol{f}(\boldsymbol{c} - \boldsymbol{b}) = (\|\boldsymbol{c} - 2\boldsymbol{b}\|, 1)$ the y-coordinate increases strictly, and the x-coordinate also increases strictly at least in a neighbourhood of $(1,0)$ if $\|\boldsymbol{c} - 2\boldsymbol{b}\| > 1$. However, by the triangle inequality we have $\|\boldsymbol{c} - 2\boldsymbol{b}\| \geq 1$, and if $\|\boldsymbol{c} - 2\boldsymbol{b}\| = 1$ then we already have found $(1,1)$ to be in the range of $\boldsymbol{f}$, which we assumed not to happen.

From $\boldsymbol{f}(\boldsymbol{c} - \boldsymbol{b})$ to $\boldsymbol{f}(-\boldsymbol{b}) = (2, \|\boldsymbol{b} + \boldsymbol{c}\|)$ the x-coordinate increases, and from $\boldsymbol{f}(-\boldsymbol{b})$ to $\boldsymbol{f}(-\boldsymbol{c}) = (\|\boldsymbol{b}+\boldsymbol{c}\|, 2)$ the y-coordinate increases. From $\boldsymbol{f}(-\boldsymbol{c})$ to $\boldsymbol{f}(\boldsymbol{b}-\boldsymbol{c}) = (1, \|\boldsymbol{b}-2\boldsymbol{c}\|)$ the x-coordinate decreases strictly, and the y-coordinate decreases. Again by the triangle inequality and assumption we have $\|\boldsymbol{b} - 2\boldsymbol{c}\| > 1$. Finally, from $\boldsymbol{f}(\boldsymbol{b} - 2\boldsymbol{c})$ to $\boldsymbol{f}(\boldsymbol{b})$ the x-coordinate decreases strictly, and the y-coordinate decreases.

It follows that $\boldsymbol{f}(S_2)$ encircles $(1,1)$, and the theorem is proved. $\qquad\square$

Note that if the fourth point which extends the equilateral set is in the plane X_2 (i.e. if $(1,1) \in \boldsymbol{f}(S_2)$ in the above proof), then X_2 is isometric to ℓ_∞^2, i.e. S_2 is a parallelogram (by Corollary 1). However, it is impossible for all 2-dimensional sections of the unit ball through the origin to be parallelograms. Thus we may always choose an X_2 so that the fourth point is outside X_2 to obtain a non-coplanar equilateral set of four points.

One would hope that ideas similar to the above proof would help to prove Conjecture 2. However, the above proof cannot be naïvely generalized, as the following example of Petty [29] shows.

Define the following norm on $\mathbb{R}^n$:

$$\|(x_1, \ldots, x_n)\| := |x_1| + \sqrt{x_2^2 + \cdots + x_n^2}.$$

Thus $\|\cdot\|$ is the ℓ_1-sum of $\mathbb{R}$ and ℓ_2^{n-1}: $(\mathbb{R}^n, \|\cdot\|) = \mathbb{R} \oplus_1 \ell_2^{n-1}$. Its unit ball is a double cone over an $(n-1)$-dimensional Euclidean ball:

$$B_{\|\cdot\|} = \text{conv}(B_2^{n-1} \cup \{\pm \boldsymbol{e}_1\}).$$

For any $n \geq 2$, $(\mathbb{R}^n, \|\cdot\|)$ contains a maximal equilateral set of four points. In fact the largest size of an equilateral set containing $\pm \boldsymbol{e}_1$ is 4. First of all note that

$$\|\boldsymbol{x} - \boldsymbol{e}_1\| = \|\boldsymbol{x} + \boldsymbol{e}_1\| \iff x_1 = 0.$$

Secondly, since $\|\boldsymbol{e}_1 - (-\boldsymbol{e}_1)\| = 2$, we want $\|\boldsymbol{x} \pm \boldsymbol{e}_1\| = 2$. Since then $x_1 = 0$, we obtain that $1 + \sqrt{x_2^2 + \cdots + x_n^2} = 2$ must hold, which gives that $\boldsymbol{x} \in \mathbb{S}^{n-2}$. Clearly there are at most two points on the Euclidean $(n-2)$-dimensional unit sphere at distance 2 (and two such points must be antipodal).

On the other hand any n equilateral points in the $(n-1)$-dimensional Euclidean subspace $x_1 = 0$ can always be extended to $n + 1$ equilateral points. For example, if we choose the n equilateral points on $\mathbb{S}^{n-2}$ then $\lambda \boldsymbol{e}_1$ will be the $(n+1)$st point, for some appropriate $\lambda \in \mathbb{R}$.

Problem 5. *Calculate $e(X)$ for the Minkowski space in the above example.*

3.2 ℓ_p^n with $1 < p < 2$

We sketch the proof that $e(\ell_p^n) > n + 1$ for any $1 < p < 2$ with n sufficiently large (depending on p).

Note that the four ± 1-vectors in $\mathbb{R}^3$ with an even number of -1's, i.e.,

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}, \tag{3}$$

is an equilateral set for any p-norm. Thus if we consider $\mathbb{R}^6$ to be the direct sum $\mathbb{R}^3 \oplus \mathbb{R}^3$ and put the above four vectors in each copy of $\mathbb{R}^3$, we obtain the following 8 vectors in $\mathbb{R}^6$:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \\ -1 \\ 1 \end{bmatrix}. \tag{4}$$

Then it is easily seen that these 8 vectors are equilateral iff $2 \cdot 2^p = 6 \iff p = \log 3 / \log 2$.

The combinatorial reason why this worked is that any two of the vectors in (3) differ in exactly two positions. For a generalization we use Hadamard matrices. An *Hadamard matrix* is an $n \times n$ matrix H with all entries ± 1 and with orthogonal columns. By multiplying some of the columns of an Hadamard matrix one obtains an Hadamard matrix with the first row containing only 1's. If we remove this first row, we obtain n column vectors in $\mathbb{R}^{n-1}$ with ± 1 entries such that any two differ in exactly $n/2$ positions. The vectors in (3) was obtained in this way from a 4×4 Hadamard matrix. We may now construct $2n$ vectors in $\mathbb{R}^{2n-2} = \mathbb{R}^{n-1} \oplus \mathbb{R}^{n-1}$ as before that will be equilateral iff

$$\frac{n}{2} 2^p = 2n - 2 \iff p = 2 + \lg \frac{n-1}{n}.$$

The only problem remaining is to find Hadamard matrices. It is easy to see that for an $n \times n$ Hadamard matrix to exist, n must be divisible by 4. It is an open problem whether the converse is true. However, the following well-known construction gives a Hadamard matrix of order 2^k:

$$H_0 = \begin{bmatrix} 1 \end{bmatrix}, \quad H_{k+1} = \begin{bmatrix} H_k & H_k \\ H_k & -H_k \end{bmatrix}, \ k \geq 0.$$

Thus we have $e(\ell_p^n) > n + 1$ for p arbitrarily close to 2.

With a more involved construction the following can be shown:

Theorem 4. *For any $1 < p < 2$ and $n \geq 1$, let*

$$k = \left\lceil \frac{\log(1 - 2^{p-2})^{-1}}{\log 2} \right\rceil - 1.$$

Then

$$e(\ell_p^n) \geq \left\lfloor \frac{2^{k+1}}{2^{k+1} - 1} n \right\rfloor.$$

In particular, if $n \geq 2^{k+2} - 2$ then $e(\ell_p^n) > n + 1$.

Corollary 2. $e(\ell_p^n) > n + 1$ *if* $p < 2 - \frac{1 + o(1)}{(\log 2)n}$.

Compare this with Corollary 5 in Section 4.2. The smallest dimension for which the theorem gives an example of $e(\ell_p^n) > n + 1$ is $n = 6$ (and $1 < p \leq \log 3 / \log 2$). In [37] there is also a 4-dimensional example (with $1 < p \leq \log \frac{5}{2} / \log 2$).

In the next two sections we show that $e(X)$ goes to infinity with the dimension of X (Brass-Dekster), although the best known lower bound for $e(X)$ is much smaller than linear in n. There are three ingredients to the proof: The Cayley-Menger theory of the embedding of metric spaces into Euclidean space, Dvoretzky's theorem, and the Brouwer fixed point theorem.

4 Cayley-Menger Theory

In this section we discuss a fragment of the Cayley-Menger theory by giving necessary and sufficient conditions for a metric space of size $n+1$ to be embeddable as an affinely independent set in ℓ_2^n. The general theory is by Menger [27]; see also Blumenthal [8, §40].

4.1 The Cayley-Menger determinant

Let a metric space on $n+1$ points $\boldsymbol{p}_0, \ldots, \boldsymbol{p}_n$ be given, with distances $\rho_{ij} = d(\boldsymbol{p}_i, \boldsymbol{p}_j)$. We now derive a necessary condition for $\boldsymbol{p}_0, \ldots, \boldsymbol{p}_n$ to be isometric to an affinely independent subset of ℓ_2^n, in terms of ρ_{ij}.

So we assume that $\{\boldsymbol{p}_0, \ldots, \boldsymbol{p}_n\}$ is an affinely independent set in ℓ_2^n with $\|\boldsymbol{p}_i - \boldsymbol{p}_j\|_2 = \rho_{ij}$. We write the coordinates of $\boldsymbol{p}_i$ as $(p_i^{(1)}, p_i^{(2)}, \ldots, p_i^{(n)})$. Let $\Delta = \mathrm{conv}\{\boldsymbol{p}_0, \ldots, \boldsymbol{p}_n\}$ denote the simplex with vertices $\boldsymbol{p}_i$. Then

$$
\pm \mathrm{vol}(\Delta) \;=\; \frac{1}{n!}
\begin{vmatrix}
p_0^{(1)} & p_1^{(1)} & \cdots & p_n^{(1)} \\
p_0^{(2)} & p_1^{(2)} & \cdots & p_n^{(2)} \\
\vdots & \vdots & \cdots & \vdots \\
p_0^{(n)} & p_1^{(n)} & \cdots & p_n^{(n)} \\
1 & 1 & \cdots & 1
\end{vmatrix}
$$

$$
= \frac{1}{n!}
\begin{vmatrix}
\boldsymbol{p}_0 & \boldsymbol{p}_1 & \cdots & \boldsymbol{p}_n \\
1 & 1 & \cdots & 1
\end{vmatrix}.
$$

Squaring (and adding an extra row and column),

$$
\mathrm{vol}(\Delta)^2 \;=\; \frac{1}{(n!)^2} \det
\begin{bmatrix}
\boldsymbol{p}_0^{\mathrm{tr}} & 1 & 0 \\
\boldsymbol{p}_1^{\mathrm{tr}} & 1 & 0 \\
\vdots & \vdots & \vdots \\
\boldsymbol{p}_n^{\mathrm{tr}} & 1 & 0 \\
\boldsymbol{o}^{\mathrm{tr}} & 0 & 1
\end{bmatrix}
\begin{bmatrix}
\boldsymbol{p}_0 & \cdots & \boldsymbol{p}_n & \boldsymbol{o} \\
1 & \cdots & 1 & 0 \\
0 & \cdots & 0 & 1
\end{bmatrix}
$$

$$
= -\frac{1}{(n!)^2} \det
\begin{bmatrix}
\boldsymbol{p}_0^{\mathrm{tr}} & 0 & 1 \\
\boldsymbol{p}_1^{\mathrm{tr}} & 0 & 1 \\
\vdots & \vdots & \vdots \\
\boldsymbol{p}_n^{\mathrm{tr}} & 0 & 1 \\
\boldsymbol{o}^{\mathrm{tr}} & 1 & 0
\end{bmatrix}
\begin{bmatrix}
\boldsymbol{p}_0 & \cdots & \boldsymbol{p}_n & \boldsymbol{o} \\
1 & \cdots & 1 & 0 \\
0 & \cdots & 0 & 1
\end{bmatrix}
\qquad \text{(interchange last two}
$$
$$
\text{columns of first matrix)}
$$

$$
= -\frac{1}{(n!)^2}
\begin{vmatrix}
\langle \boldsymbol{p}_0, \boldsymbol{p}_0 \rangle & \langle \boldsymbol{p}_0, \boldsymbol{p}_1 \rangle & \cdots & \langle \boldsymbol{p}_0, \boldsymbol{p}_n \rangle & 1 \\
\langle \boldsymbol{p}_1, \boldsymbol{p}_0 \rangle & \langle \boldsymbol{p}_1, \boldsymbol{p}_1 \rangle & \cdots & \langle \boldsymbol{p}_1, \boldsymbol{p}_n \rangle & 1 \\
\vdots & \vdots & \cdots & \vdots & \\
\langle \boldsymbol{p}_n, \boldsymbol{p}_0 \rangle & \langle \boldsymbol{p}_n, \boldsymbol{p}_1 \rangle & \cdots & \langle \boldsymbol{p}_n, \boldsymbol{p}_n \rangle & 1 \\
1 & 1 & \cdots & 1 & 0
\end{vmatrix}
$$

$$
= -\frac{1}{2^{n+1}(n!)^2}
\begin{vmatrix}
0 & -\rho_{01}^2 & \cdots & -\rho_{0n}^2 & 1 \\
-\rho_{10}^2 & 0 & \ddots & \vdots & \vdots \\
\vdots & \ddots & \ddots & \vdots & \vdots \\
-\rho_{n0}^2 & \cdots & -\rho_{n,n-1}^2 & 0 & 1 \\
1 & \cdots & \cdots & \cdots & 0
\end{vmatrix}
\qquad
\begin{array}{l}
\text{(use } 2\langle \boldsymbol{p}_i, \boldsymbol{p}_j \rangle = \\
\langle \boldsymbol{p}_i, \boldsymbol{p}_i \rangle + \langle \boldsymbol{p}_j, \boldsymbol{p}_j \rangle - \rho_{ij}^2 \text{ and} \\
\text{subtract multiples of the} \\
\text{last row and column)}
\end{array}
$$

$$
= \frac{(-1)^{n+1}}{2^{n+1}(n!)^2} \det
\begin{bmatrix}
& & & 1 \\
& P & & \vdots \\
& & & 1 \\
1 & \cdots & 1 & 0
\end{bmatrix},
$$

where $P = \left[\rho_{ij}^2\right]$. We call

$$\mathrm{CMdet}(\boldsymbol{p}_0, \ldots, \boldsymbol{p}_n) := \begin{vmatrix} & & & 1 \\ & P & & \vdots \\ & & & 1 \\ 1 & \cdots & 1 & 0 \end{vmatrix}$$

the *Cayley-Menger determinant* of the $(n+1)$-point metric space. We have shown the following.

Proposition 2. *If a metric space on $n+1$ points can be embedded into ℓ_2^n as an affinely independent set, then its Cayley-Menger determinant has sign $(-1)^{n+1}$. (Also, if the metric space can be embedded as an affinely dependent set, then its Cayley-Menger determinant is 0.)*

Conversely we have

Theorem 5 (Menger [27]). *A metric space on $n+1$ points $\boldsymbol{p}_0, \ldots, \boldsymbol{p}_n$ can be embedded into ℓ_2^n as an affinely independent set if for each $k = 1, \ldots, n$, the Cayley-Menger determinant of $\boldsymbol{p}_0, \ldots, \boldsymbol{p}_k$ has sign $(-1)^{k+1}$.*

Proof. For $n = 1$,

$$\mathrm{CMdet}(\boldsymbol{p}_0, \boldsymbol{p}_1) = \begin{vmatrix} 0 & \rho_{01}^2 & 1 \\ \rho_{10}^2 & 0 & 1 \\ 1 & 1 & 0 \end{vmatrix} = \rho_{01}^2 + \rho_{10}^2 > 0.$$

Thus the theorem is trivial for $n = 1$ (and even more so for $n = 0$).

Thus we may assume as inductive hypothesis that the theorem is true for $n - 1$. Thus $\boldsymbol{p}_0, \ldots, \boldsymbol{p}_{n-1}$ can be embedded into $\ell_2^{n-1} \subset \ell_2^n$ as an affinely independent set, say $\boldsymbol{p}_0 \mapsto \boldsymbol{o}$ and $\boldsymbol{p}_i \mapsto \boldsymbol{x}_i$, $i = 1, \ldots, n-1$. Then $\mathrm{span}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1}\} = \ell_2^{n-1}$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1}$ are linearly independent.

We have $\|\boldsymbol{x}_i\|_2 = \rho_{i0}$, and $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2 = \rho_{ij}$ for $1 \leq i, j \leq n-1$, and we have to find an $\boldsymbol{x} \notin \ell_2^{n-1}$ satisfying

$$\|\boldsymbol{x}\|_2 = \rho_{0n} \tag{5}$$

and

$$\|\boldsymbol{x} - \boldsymbol{x}_i\|_2 = \rho_{in} \text{ for all } i = 1, \ldots, n-1. \tag{6}$$

We write $\boldsymbol{x} = \boldsymbol{v} + \lambda \boldsymbol{e}_n$ uniquely, where $\boldsymbol{v} \in \ell_2^{n-1}$ and $\lambda \in \mathbb{R}$. If we square (6) and simplify, we find that $\boldsymbol{v}$ has to satisfy the following $n - 1$ linear equations in ℓ_2^{n-1}:

$$2 \langle \boldsymbol{x}_i, \boldsymbol{v} \rangle = \|\boldsymbol{x}\|_2^2 + \rho_{0i}^2 - \rho_{in}^2, \; i = 1, \ldots, n-1.$$

Since the $\boldsymbol{x}_i$ are linearly independent, there is a unique solution $\boldsymbol{v} \in \ell_2^{n-1}$ to the following modified equation:

$$2 \langle \boldsymbol{x}_i, \boldsymbol{v} \rangle = \rho_{0n}^2 + \rho_{0i}^2 - \rho_{in}^2, \; i = 1, \ldots, n-1, \tag{7}$$

and we will have found the required $\boldsymbol{x}$ once we can satisfy (5), which is equivalent to

$$\lambda^2 = \rho_{0n}^2 - \|\boldsymbol{v}\|_2^2.$$

Thus it remains to prove that $\rho_{0n}^2 - \|\boldsymbol{v}\|_2^2 > 0$. To this end we calculate the Cayley-Menger determinant of $\boldsymbol{o}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1}, \boldsymbol{v}$, which is 0 by Proposition 2:

$$\begin{vmatrix} 0 & \rho_{01}^2 & \cdots & \rho_{0,n-1}^2 & \|\boldsymbol{v}\|_2^2 & 1 \\ \rho_{10}^2 & 0 & \cdots & \rho_{1,n-1}^2 & \|\boldsymbol{v} - \boldsymbol{x}_1\|_2^2 & 1 \\ \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \rho_{n-1,0}^2 & \cdots & \cdots & 0 & \|\boldsymbol{v} - \boldsymbol{x}_{n-1}\|_2^2 & 1 \\ \|\boldsymbol{v}\|_2^2 & \|\boldsymbol{v} - \boldsymbol{x}_1\|_2^2 & \cdots & \|\boldsymbol{v} - \boldsymbol{x}_{n-1}\|_2^2 & 0 & 1 \\ 1 & 1 & \cdots & 1 & 1 & 0 \end{vmatrix} = 0.$$

Multiplying out

$$\|\boldsymbol{v} - \boldsymbol{x}_i\|_2^2 = \|\boldsymbol{v}\|_2^2 - 2\langle \boldsymbol{x}_i, \boldsymbol{v}\rangle + \|\boldsymbol{x}_i\|_2^2$$
$$= \|\boldsymbol{v}\|_2^2 - \rho_{0n}^2 + \rho_{in}^2 \quad \text{(by (7))},$$

and using the last row and column to eliminate $\|\boldsymbol{v}\|_2^2 - \rho_{0n}^2$, we obtain

$$\begin{vmatrix} 0 & \rho_{01}^2 & \cdots & \rho_{0,n-1}^2 & \rho_{0n}^2 & 1 \\ \rho_{10}^2 & 0 & \cdots & \rho_{1,n-1}^2 & \rho_{1n}^2 & 1 \\ \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \rho_{n-1,0}^2 & \cdots & \cdots & 0 & \rho_{n-1,n}^2 & 1 \\ \rho_{n,0}^2 & \rho_{n,1}^2 & \cdots & \rho_{n,n-1}^2 & -2(\|\boldsymbol{v}\|_2^2 - \rho_{0n}^2) & 1 \\ 1 & 1 & \cdots & 1 & 1 & 0 \end{vmatrix} = 0.$$

This determinant differs from $\text{CMdet}(\boldsymbol{p}_0, \ldots, \boldsymbol{p}_n)$ only in the second last column. Thus we may subtract this determinant from $\text{CMdet}(\boldsymbol{p}_0, \ldots, \boldsymbol{p}_n)$, which we are given has sign $(-1)^{n+1}$, to obtain the following determinant of sign $(-1)^{n+1}$:

$$\begin{vmatrix} 0 & \rho_{01}^2 & \cdots & \rho_{0,n-1}^2 & 0 & 1 \\ \rho_{10}^2 & 0 & \cdots & \rho_{1,n-1}^2 & 0 & 1 \\ \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \rho_{n-1,0}^2 & \cdots & \cdots & 0 & 0 & 1 \\ \rho_{n,0}^2 & \rho_{n,1}^2 & \cdots & \rho_{n,n-1}^2 & 2(\|\boldsymbol{v}\|_2^2 - \rho_{0n}^2) & 1 \\ 1 & 1 & \cdots & 1 & 1 & 0 \end{vmatrix}$$
$$= 2(\|\boldsymbol{v}\|_2^2 - \rho_{0n}^2)\,\text{CMdet}(\boldsymbol{p}_0, \ldots, \boldsymbol{p}_{n-1}) \quad \text{(expanding along 2nd last column)}.$$

Since it is also given that $\text{CMdet}(\boldsymbol{p}_0, \ldots, \boldsymbol{p}_{n-1})$ has sign $(-1)^n$, it follows that $\|\boldsymbol{v}\|_2^2 - \rho_{0n}^2 < 0$. This gives us a $\lambda \in \mathbb{R}$ such that $x = \boldsymbol{v} + \lambda \boldsymbol{e}_n$ satisfies (5) as well. The theorem is proved. $\qquad\square$

4.2 Embedding an almost equilateral n-simplex into ℓ_2^n

We now want to use Theorem 5 to show that an n-simplex with all side lengths close to 1 can be embedded isometrically into ℓ_2^n. Thus we have to find the sign of the determinant of a matrix with 0 on the diagonal and all off-diagonal entries close to 1, since this is how the Cayley-Menger determinant will look like.

We start with the following simple observations. Let J be the $n \times n$ matrix with all entries equal to 1.

Lemma 6. *Let $n \geq 1$.*

1. *The characteristic polynomial of J is $(-1)^n(x-n)x^{n-1}$.*

2. *Therefore the characteristic polynomial of $J - I$ is $(-1)^n(x-n+1)(x+1)^{n-1}$, and $\det(J-I) = (-1)^{n-1}(n-1)$.*

3. *Therefore $J-I$ is invertible, and its inverse has eigenvalues -1 with multiplicity $n-1$, and $1/(n-1)$.*

Proposition 3. *Let $A = \begin{bmatrix} a_{ij} \end{bmatrix}$ be an $(n+2) \times (n+2)$ matrix with $a_{ii} = 0$, $|1 - a_{ij}| \leq 1/(n - \frac{3}{2})$ for $1 \leq i \neq j \leq n+1$, and $a_{i,n+2} = a_{n+2,i} = 1$ for $i = 1, \ldots, n$. Then $\det(A)$ has sign $(-1)^{n+1}$.*

Proof. We first prove that A is invertible. Let $E = \begin{bmatrix} \varepsilon_{ij} \end{bmatrix} := J - I - A$. Then $\varepsilon_{ii} = 0$ and $\varepsilon_{i,n+2} = \varepsilon_{n+2,i} = 0$ for all i, and $|\varepsilon_{ij}| \leq 1/(n - \frac{3}{2})$ for all $1 \leq i \neq j \leq n+1$. Since $A = J - I - E = (J-I)(I - (J-I)^{-1}E)$, it is sufficient to prove that $\|(J-I)^{-1}E\|_2 < 1$, where $\|\cdot\|_2$ is the operator norm on $L(\ell_2^n \to \ell_2^n)$. Firstly, since the operator norm is the maximum modulus of the eigenvalues, we have by Lemma 6 that

12

$\|(J-I)^{-1}\|_2 = 1$. Secondly, by the Cauchy-Schwartz inequality, $\|E\|_2 \leq \sqrt{\sum_{i,j=1}^{n+2} \varepsilon_{ij}^2} < 1$. Thus A is invertible.

To show that $\det A$ has sign $(-1)^{n+1}$, observe that we may join $J - I$ to A with a curve A_t, $0 \leq t \leq 1$, where each A_t satisfies the hypothesis of the proposition. Thus we know from what we have just proved that each A_t is invertible. By continuity considerations the sign of $\det A_t$ does not change. Therefore, A and $J - I$ have the same sign, which is $(-1)^{n+1}$ by Lemma 6. $\square$

Theorem 6 (Dekster & Wilker [14]). *Let $\boldsymbol{p}_0, \ldots, \boldsymbol{p}_n$ be a metric space with distances satisfying*

$$1 \leq d(\boldsymbol{p}_i, \boldsymbol{p}_j) \leq 1 + \frac{1}{n+1}.$$

Then $\boldsymbol{p}_0, \ldots, \boldsymbol{p}_n$ can be embedded into ℓ_2^n, and any such embedding must be an affinely independent set.

Proof. We first scale the distances by some $\alpha > 0$ in such a way that the given inequality is transformed into

$$1 - \varepsilon \leq (\alpha d(\boldsymbol{p}_i, \boldsymbol{p}_j))^2 \leq 1 + \varepsilon$$

for some $\varepsilon > 0$. Thus ε and α must satisfy

$$\frac{1-\varepsilon}{\alpha^2} = 1 \text{ and } \frac{1+\varepsilon}{\alpha^2} = \left(\frac{n+2}{n+1}\right)^2.$$

Eliminating α we find

$$\frac{1+\varepsilon}{1-\varepsilon} = \left(\frac{n+2}{n+1}\right)^2. \tag{8}$$

We want to apply Theorem 5. Since the CM determinant of $\boldsymbol{p}_0, \ldots, \boldsymbol{p}_k$ is a $(k+2) \times (k+2)$ matrix with $k \leq n$, to show that it has sign $(-1)^{k+1}$, it is sufficient to prove that

$$\varepsilon \leq \frac{1}{n + \frac{3}{2}},$$

according to Proposition 3. By (8) we have

$$
\begin{aligned}
\varepsilon &= \frac{\left(\frac{n+2}{n+1}\right)^2 - 1}{\left(\frac{n+2}{n+1}\right)^2 + 1} \\
&= \frac{(n+2)^2 - (n+1)^2}{(n+2)^2 + (n+1)^2} = \frac{2n+3}{2n^2 + 6n + 5} \\
&< \frac{2n+3}{2n^2 + 6n + \frac{9}{2}} = \frac{2}{2n+3}.
\end{aligned}
$$

By Proposition 2 no embedding can be affinely dependent. $\square$

Dekster and Wilker found the smallest $\gamma_n > 1$ such that whenever

$$1 \leq d(\boldsymbol{p}_i, \boldsymbol{p}_j) \leq \gamma_n$$

then $\boldsymbol{p}_0, \ldots, \boldsymbol{p}_{n+1}$ can be embedded into ℓ_2^n: it is

$$\gamma_n = \begin{cases} \sqrt{1 + \frac{2n+2}{n^2-2}} & \text{if } n \text{ is even,} \\ \sqrt{1 + \frac{2}{n-1}} & \text{if } n \text{ is odd.} \end{cases}$$

Their proof is geometrical. The above analytical proof gives values that differ from γ_n by $O(n^{-2})$ as $n \to \infty$.

Corollary 3 (Schütte [30]). *For any $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{n+2} \in \ell_2^n$ we have*

$$\frac{\max\|\boldsymbol{p}_i - \boldsymbol{p}_j\|_2}{\min_{i \neq j}\|\boldsymbol{p}_i - \boldsymbol{p}_j\|_2} > 1 + \frac{1}{n+2}.$$

Schütte in fact determined the smallest $\delta_n > 1$ such that

$$\frac{\max\|\boldsymbol{p}_i - \boldsymbol{p}_j\|_2}{\min_{i \neq j}\|\boldsymbol{p}_i - \boldsymbol{p}_j\|_2} \geq 1 + \delta_n$$

for all possible $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{n+2} \in \ell_2^n$: it is $\delta_n = \gamma_{n+1}$. This result was also discovered by Seidel [31], and follows from the embedding result of Dekster and Wilker [14]; see [4] for a very simple proof.

Corollary 4. *If $d(X, \ell_2^n) \leq 1 + \frac{1}{n+2}$, then $e(X) \leq n + 1$.*

Corollary 5 (Smyth [33]). *If $|p - 2| < \frac{4(1+o(1))}{n \log n}$ then $e(\ell_p^n) = n + 1$.*

Compare this with Corollary 2 in Section 3.2.

5   The Theorem of Brass and Dekster

Theorem 7 (Brass [10] & Dekster [13]). *An n-dimensional Minkowski space contains an equilateral set of size $c(\log n)^{1/3}$ for some constant $c > 0$ and n sufficiently large.*

Both Brass' and Dekster's proofs use Dvoretzky's theorem combined with a topological result. We follow Brass' proof, which uses only the Brouwer fixed point theorem.

For the proof we need Dvoretzky's theorem.

Dvoretzky's Theorem. *There exists a contant $c > 0$ such that for any $\varepsilon > 0$, any Minkowski space X of sufficiently large dimension contains an m-dimensional subspace with Banach-Mazur distance at most $1 + \varepsilon$ to ℓ_2^m. We may take $\dim X \geq e^{cm\varepsilon^{-2}}$ for some absolute constant $c > 0$.*

The estimate for $\dim X$, which is best possible (up to the value of c), was proven by Gordon [16].

Proof of Theorem 7. Denote the n-dimensional space by X. Let $m = c_1(\log n)^{1/3}$ and $\varepsilon = 1/(m+1)$. A simple calculation gives that $e^{cm\varepsilon^{-2}} < n$ if we choose $c_1 = c_1(c)$ sufficiently small. Dvoretzky's theorem gives that X contains an m-dimensional subspace Y with Banach-Mazur distance at most $1 + \frac{1}{m+1}$ to ℓ_2^m. The proof is then finished by applying the following theorem. $\square$

Theorem 8 (Brass [10] & Dekster [13]). *Let X be an n-dimensional Minkowski space with Banach-Mazur distance $d(X, \ell_2^n) \leq 1 + \frac{1}{n+1}$. Then an equilateral set in X of at most n points can be extended to one of $n + 1$ points. In particular, $e(X) \geq n + 1$.*

Proof. By induction on nested subspaces of X it is sufficient to prove that a 1-equilateral set $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$ may be extended. We may assume that a coordinate system has been chosen such that

$$\|\boldsymbol{x}\| \leq \|\boldsymbol{x}\|_2 \leq (1 + \varepsilon)\|\boldsymbol{x}\|,$$

where $\varepsilon = (n+1)^{-1}$. Let $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$ be a 1-equilateral set in X. We want to find a $\boldsymbol{p} \in X$ such that $\|\boldsymbol{p} - \boldsymbol{p}_i\| = 1$ for all $i = 1, \ldots, n$.

We have

$$1 \leq \|\boldsymbol{p}_i - \boldsymbol{p}_j\|_2 \leq 1 + \varepsilon \text{ for all distinct } i, j.$$

By Theorem 6 $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$ are affinely independent. For any $\rho_1, \ldots, \rho_n \in [1, 1 + \varepsilon]$ there exists a point $\boldsymbol{x} \in X$ with $\|\boldsymbol{x} - \boldsymbol{p}_i\|_2 = \rho_i$, $i = 1, \ldots, n$, by Theorem 6. Furthermore, by fixing a half space bounded by the hyperplane through $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$, there is a unique such $\boldsymbol{x}$ in this half space (by simple linear

algebra arguments as in the proof of of Theorem 5; exercise). We write $\boldsymbol{x} =: \boldsymbol{f}(\rho_1, \ldots, \rho_n)$, and define $\varphi : [0, \varepsilon]^n \to [0, \varepsilon]^n$ by sending $(\delta_1, \ldots, \delta_n) \mapsto (y_1, \ldots, y_n)$, where

$$y_i = \delta_i + 1 - \|\boldsymbol{f}(1 + \delta_1, \ldots, 1 + \delta_n) - \boldsymbol{p}_i\|, \ i = 1, \ldots, n.$$

Since $1 + \delta_i \leq 1 + \varepsilon$, each y_i is well-defined. Also,

$$
\begin{aligned}
y_i \ &\leq \ \delta_i + 1 - (1 + \varepsilon)^{-1}\|\boldsymbol{f}(1 + \delta_1, \ldots, 1 + \delta_n) - \boldsymbol{p}_i\|_2 \\
&= \ \delta_i + 1 - (1 + \varepsilon)^{-1}(1 + \delta_i) \\
&= \ (1 - (1 + \varepsilon)^{-1})(1 + \delta_i) \\
&\leq \ (1 - (1 + \varepsilon)^{-1})(1 + \varepsilon) = \varepsilon,
\end{aligned}
$$

and

$$
\begin{aligned}
y_i \ &\geq \ \delta_i + 1 - \|\boldsymbol{f}(1 + \delta_1, \ldots, 1 + \delta_n) - \boldsymbol{p}_i\|_2 \\
&= \ \delta_i + 1 - (1 + \delta_i) = 0,
\end{aligned}
$$

which shows that φ maps into $[0, \varepsilon]^n$.

By Brouwer's fixed point theorem, φ has a fixed point $(\delta_1, \ldots, \delta_n)$, which gives

$$\|\boldsymbol{f}(1 + \delta_1, \ldots, 1 + \delta_n) - \boldsymbol{p}_i\| = 1 \ \text{ for all } i = 1, \ldots, n. \qquad \square$$

Corollary 6 (Brass [10] & Dekster [13]). *If $d(X, \ell_2^n) \leq 1 + \frac{1}{n+2}$ then $e(X) = n + 1$.*

Proof. Combine the above theorem with Corollary 4. $\qquad \square$

6   The Linear Algebra Method

Linear algebra provides a very powerful counting tool. The basic idea is that whenever one has a set S of m elements, and one wants to find an upper bound to m, one constructs a vector $\boldsymbol{v}_s$ in some N-dimensional vector space for each element of $s \in S$. If one can prove that the vectors constructed are linearly independent, then one immediately has the upper bound $m \leq N$. If the vectors are not linearly independent, then one may try to prove that the square matrix $A^{\mathrm{tr}}A$ has a high rank, where A is the matrix with column vectors $\boldsymbol{v}_s, s \in S$.

6.1   Linear independence

As a first example, one can prove that $e(\ell_2^n) \leq n + 1$ as follows. Let $S \subset \ell_2^n$ be 1-equilateral, and define for each $\boldsymbol{s} \in S$ a polynomial in n variables

$$P_{\boldsymbol{s}}(\boldsymbol{x}) = P_{\boldsymbol{s}}(x_1, \ldots, x_n) := 1 - \|\boldsymbol{x} - \boldsymbol{s}\|_2^2.$$

Then it is easily seen that $\{1\} \cup \{P_{\boldsymbol{s}}(\boldsymbol{x}) : \boldsymbol{s} \in S\}$ is linearly independent in the vector space $\mathbb{R}[x_1, \ldots, x_n]$ of polynomials in $x_1, \ldots, x_n$ with real coefficients, and secondly that all $P_{\boldsymbol{s}}(\boldsymbol{x})$ are in the linear span of $\{1, \sum_{i=1}^n x_i^2, x_1, \ldots, x_n\}$. It follows that $1 + |S| \leq 2 + n$, which is what was required.

In the same way one may prove that $e(\ell_p^n) \leq 1 + (p - 1)n$ if p is an even integer (an observation due to Galvin; see [33]). With some more work the following may be shown:

Theorem 9 (Swanepoel [37]). *For each $n \geq 1$, $e(\ell_4^n) = n + 1$.*

Proof. The lower bound has already been shown in Section 1.2.

Let S be a 1-equilateral set in ℓ_4^n. For each $\boldsymbol{s} \in S$, let $P_{\boldsymbol{s}}(x) = p_{\boldsymbol{s}}(x_1, \ldots, x_n)$ be the following polynomial:

$$
\begin{aligned}
P_{\boldsymbol{s}}(\boldsymbol{x}) \ &:= \ 1 - \|\boldsymbol{x} - \boldsymbol{s}\|_4^4 \\
&= \ 1 - \|\boldsymbol{s}\|_4^4 - \sum_{i=1}^n x_i^4 + 4 \sum_{i=1}^n s_i x_i^3 - 6 \sum_{i=1}^n s_i^2 x_i^2 + 4 \sum_{i=1}^n s_i^3 x_i. \qquad (9)
\end{aligned}
$$

15

Thus each $P_{\boldsymbol{s}}$ is in the linear span of

$$\{1, \sum_{i=1}^{n} x_i^4\} \cup \{x_i^k : i = 1, \dots, n; k = 1, 2, 3\},$$

which is a subspace of $\mathbb{R}[x_1, \dots, x_n]$ of dimension $2 + 3n$. Since $P_{\boldsymbol{s}}(\boldsymbol{s}) = 1$ and $P_{\boldsymbol{s}}(\boldsymbol{s}') = 0$ for all distinct $\boldsymbol{s}, \boldsymbol{s}' \in S$ we have that $\{P_{\boldsymbol{s}} : \boldsymbol{s} \in S\}$ is linearly independent. (So already $|S| \le 2 + 3n$.)

We now prove that

$$\{P_{\boldsymbol{s}} : \boldsymbol{s} \in S\} \cup \{1\} \cup \{x_i^k : i = 1, \dots, n; k = 1, 2\}$$

is still linearly independent, which will give $|S| + 1 + 2n \le 2 + 3n$, proving the theorem.

Let $\lambda_{\boldsymbol{s}} (\boldsymbol{s} \in S), \lambda, \lambda_i, \mu_i (i = 1, \dots, d) \in \mathbb{R}$ satisfy

$$\sum_{\boldsymbol{s} \in S} \lambda_{\boldsymbol{s}} P_{\boldsymbol{s}}(\boldsymbol{x}) + \lambda 1 + \sum_{i=1}^{n} \lambda_i x_i + \sum_{i=1}^{n} \mu_i x_i^2 = 0 \qquad \forall x_1, \dots, x_n \in \mathbb{R}. \tag{10}$$

Substitute (9) into (10):

$$(-\sum_{\boldsymbol{s} \in S} \lambda_{\boldsymbol{s}}) \sum_{i=1}^{n} x_i^4 + 4 \sum_{i=1}^{n} (\sum_{\boldsymbol{s} \in S} \lambda_{\boldsymbol{s}} s_i) x_i^3 + \sum_{i=1}^{n} (\mu_i - 6 \sum_{\boldsymbol{s} \in S} \lambda_{\boldsymbol{s}} s_i^2) x_i^2$$

$$+ \sum_{i=1}^{n} (\lambda_i + 4 \sum_{\boldsymbol{s} \in S} s_i^3) x_i + \sum_{\boldsymbol{s} \in S} \lambda_{\boldsymbol{s}} - \sum_{\boldsymbol{s} \in S} \lambda_{\boldsymbol{s}} \|\boldsymbol{s}\|_4^4 + \lambda = 0.$$

Since the above is a polynomial in $\mathbb{R}[x_1, \dots, x_n]$ which is identically 0, the coefficients are also 0:

$$\sum_{\boldsymbol{s} \in S} \lambda_{\boldsymbol{s}} = 0, \tag{11}$$

$$\sum_{\boldsymbol{s} \in S} \lambda_{\boldsymbol{s}} s_i = 0 \qquad \forall i = 1, \dots, n, \tag{12}$$

$$\mu_i - 6 \sum_{\boldsymbol{s} \in S} \lambda_{\boldsymbol{s}} s_i^2 = 0 \qquad \forall i = 1, \dots, n, \tag{13}$$

$$\lambda_i + 4 \sum_{\boldsymbol{s} \in S} \lambda_{\boldsymbol{s}} s_i^3 = 0 \qquad \forall i = 1, \dots, d, \tag{14}$$

$$\lambda + \sum_{\boldsymbol{s} \in S} \lambda_{\boldsymbol{s}} - \sum_{\boldsymbol{s} \in S} \lambda_{\boldsymbol{s}} \|\boldsymbol{s}\|_4^4 = 0. \tag{15}$$

By substituting $\boldsymbol{x} = \boldsymbol{s} \in S$ into (10) we obtain

$$\lambda_{\boldsymbol{s}} + \lambda + \sum_{i=1}^{n} \lambda_i s_i + \sum_{i=1}^{n} \mu_i s_i^2 = 0 \qquad \forall \boldsymbol{s} \in S. \tag{16}$$

Now multiply (16) by $\lambda_{\boldsymbol{s}}$ and sum over all $\boldsymbol{s} \in S$:

$$\sum_{\boldsymbol{s} \in S} \lambda_{\boldsymbol{s}}^2 + \lambda \sum_{\boldsymbol{s} \in S} \lambda_{\boldsymbol{s}} + \sum_{i=1}^{n} \lambda_i (\sum_{\boldsymbol{s} \in S} \lambda_{\boldsymbol{s}} s_i) + \sum_{i=1}^{n} \mu_i (\sum_{\boldsymbol{s} \in S} \lambda_{\boldsymbol{s}} s_i^2) = 0.$$

Then use (11)–(13) to simplify this expression as follows:

$$\sum_{\boldsymbol{s} \in S} \lambda_{\boldsymbol{s}}^2 + \frac{1}{6} \sum_{i=1}^{n} \mu_i^2 = 0.$$

It follows that all $\lambda_{\boldsymbol{s}} = 0$ and all $\mu_i = 0$. From (14) it follows that all $\lambda_i = 0$ and from (15) that $\lambda = 0$. Thus we have linear independence. $\qquad \square$

In the same way the following may be proved:

Theorem 10 (Swanepoel [37]). *For p an even integer and $n \geq 1$ we have*

$$e(\ell_p^n) \leq \begin{cases} (\frac{p}{2}-1)n + 1 & \text{if } p \equiv 0 \pmod 4, \\ \frac{p}{2}n + 1 & \text{if } p \equiv 2 \pmod 4. \end{cases}$$

6.2 Rank arguments

The following lemma is very simple, yet extremely powerful.

Rank Lemma. *Let $A = [a_{ij}]$ be a real symmetric $n \times n$ non-zero matrix. Then*

$$\operatorname{rank} A \geq \frac{(\sum_{i=1}^n a_{ii})^2}{\sum_{i,j=1}^n a_{ij}^2}.$$

Proof. Let $r = \operatorname{rank} A$, and let $\lambda_1, \ldots, \lambda_r \in \mathbb{R}$ be all non-zero eigenvalues of A. Then $\sum_i a_{ii} = \operatorname{trace}(A) = \sum_{i=1}^r \lambda_i$. Also, $\sum_{i,j} a_{ij}^2 = \operatorname{trace}(A^2) = \sum_{i=1}^r \lambda_i^2$, since A^2 has non-zero eigenvalues λ_i^2. Therefore,

$$\frac{(\sum_{i=1}^n a_{ii})^2}{\sum_{i,j=1}^n a_{ij}^2} = \frac{(\sum_{i=1}^r \lambda_i)^2}{\sum_{i,j=1}^r \lambda_i^2} \leq r$$

by the Cauchy-Schwartz inequality. $\qquad\square$

Corollary 7. *Let $A = [a_{ij}]$ be a real symmetric $n \times n$ matrix with $a_{ii} = 1$ for all i and $|a_{ij}| \leq \varepsilon$ for all $i \neq j$. Then*
$$\operatorname{rank} A \geq \frac{n}{1 + (n-1)\varepsilon^2}.$$

In particular, if $\varepsilon = n^{-1/2}$ then $\operatorname{rank} A \geq n/2$.

Also, if $\varepsilon < 1/(n-1)$ then the corollary gives that A is invertible, which also follows from the fact that A is then diagonally dominated.

The rank lemma together with a simple rounding argument gives an upper bound for $e(\ell_p^n)$, $1 \leq p < \infty$, that is polynomial in p and n. This is due to Alon (Smyth, personal communication).

Theorem 11. *For some constant $c > 0$ we have that for $1 \leq p < \infty$ and $n \geq 1$, $e(\ell_p^n) \leq cp^2 n^{2+2/p}$.*

Proof. Let $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_m$ be 1-equilateral in ℓ_p^n. Since $|p_i^{(k)} - p_j^{(k)}| \leq 1$ for all $i \neq j$ and all $k = 1, \ldots, m$, we may assume after a translation that all $p_i^{(k)} \in [0,1]$. Choose an integer N such that

$$(e-1)pn^{1/p}\sqrt{m} \leq N < N + 1 \leq epn^{1/p}\sqrt{m}.$$

(The interval $pn^{1/p}\sqrt{m} \geq 2$ since $m \geq 4$ without loss of generality.) We now round each $\boldsymbol{p}_i$ to a point in the lattice $\{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N}{N}\}^n$ as follows. Let $q_i^{(k)}$ be the integer multiple of $1/N$ nearest to $p_i^{(k)}$, say $q_i^{(k)} = \frac{d(i,k)-1}{N}$ where $1 \leq d(i,k) \leq N + 1$ (if there is a tie, choose arbitrarily). Let $\boldsymbol{q}_i = (q_i^{(1)}, \ldots, q_i^{(n)})$, and let Q be the $m \times m$ matrix $[1 - \|\boldsymbol{q}_i - \boldsymbol{q}_j\|_p^p]_{i,j}^m$. Thus Q is an approximation of the identity I_m, and we now estimate how close. The diagonal contains only 1's. For each k, $|p_i^{(k)} - q_i^{(k)}| \leq 1/2N$, and

$$\begin{aligned} \|\boldsymbol{q}_i - \boldsymbol{q}_j\|_p &\leq \|\boldsymbol{p}_i - \boldsymbol{p}_j\|_p + \|\boldsymbol{p}_i - \boldsymbol{q}_i\|_p + \|\boldsymbol{p}_j - \boldsymbol{q}_j\|_p \\ &\leq 1 + n^{1/p}/N, \end{aligned}$$

and similarly, $\|\boldsymbol{q}_i - \boldsymbol{q}_j\|_p \geq 1 - n^{1/p}/N$. Thus

$$\|\boldsymbol{q}_i - \boldsymbol{q}_j\|_p^p \leq \left(1 + \frac{n^{1/p}}{N}\right)^p \leq \left(1 + \frac{1}{(e-1)p\sqrt{m}}\right)^p < e^{\frac{1}{(e-1)\sqrt{m}}} < 1 + \frac{1}{\sqrt{m}},$$

17

since $e^x < 1 + (e-1)x$ for $0 < x < 1$. It follows that $1 - \|\boldsymbol{q}_i - \boldsymbol{q}_j\|_p^p > -1/\sqrt{m}$. Since $(1+x)^p + (1-x)^p > 2$ for $0 < x < 1$,

$$1 - \|\boldsymbol{q}_i - \boldsymbol{q}_j\|_p^p \leq 1 - \left(1 - \frac{n^{1/p}}{N}\right)^p < \left(1 + \frac{n^{1/p}}{N}\right)^p - 1 < \frac{1}{\sqrt{m}}.$$

Thus by Corollary 7, $\operatorname{rank} Q \geq m/2$. Note that $Q = \sum_{k=1}^n Q_k$, where

$$Q_k = \left[\tfrac{1}{n} - |q_i^{(k)} - q_j^{(k)}|^p\right]_{i,j=1}^m.$$

Define the $(N+1) \times m$ matrix

$$A_k = \begin{bmatrix} \boldsymbol{e}_{d(1,k)} & \boldsymbol{e}_{d(2,k)} & \cdots & \boldsymbol{e}_{d(m,k)} \end{bmatrix},$$

where $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{N+1}$ is the standard basis of $\mathbb{R}^{N+1}$ (recall $q_i^{(k)} = \frac{d(i,k)-1}{N}$). Secondly, define

$$B = \left[\tfrac{1}{n} - |\tfrac{i-1}{N} - \tfrac{j-1}{N}|^p\right]_{i,j=1}^{N+1}.$$

Then it is easily seen that $Q_k = A_k^{\mathrm{tr}} B A_k$. It follows that $\operatorname{rank} Q_k \leq N + 1$, hence

$$\operatorname{rank} Q \leq \sum_{k=1}^n \operatorname{rank} Q_k \leq n(N+1).$$

Putting the upper and lower bounds for $\operatorname{rank} Q$ together we obtain

$$m/2 < n(N+1) \leq epn^{1+1/p}\sqrt{m},$$

and $m < 4e^2 p^2 n^{2+2/p}$. $\qquad\square$

7 Approximation Theory: Smyth's approach

In the proof of Theorem 11 we essentially approximated the function $f_p(t) = |t|^p$, $t \in [-1,1]$ uniformly by step functions (look at the definition of the matrix B in the proof). For $p > 1$ the function f_p is differentiable $\lfloor p \rfloor$ times, so one would expect that there are better uniform approximations to f_p using polynomials or splines instead of step functions. This is indeed the case, and can be used to improve the bound of Theorem 11. Smyth [33] used the theorems of Jackson from approximation theory to find the first non-trivial upper bounds for $e(\ell_p^n)$. Alon and Pudlák [1] improved his bounds using the rank lemma. We state the theorem of Jackson that is needed (called Jackson V) and then prove their result. The proofs of the Jackson theorems may be found in many texts on approximation theory, e.g. [11].

Jackson V. *There is an absolute constant $c > 0$ such that for any $f \in C[-1,1]$*

1. *there exists a polynomial P of degree at most n such that*

   $$\|f - P\|_\infty \leq c\omega(f, 1/n),$$

   where $\omega(f, \delta) := \sup\{|f(x) - f(y)| : |x - y| \leq \delta\}$ is the modulus of continuity *of f,*

2. *and if f is k times differentiable and $n \geq k$, then there exists a polynomial P of degree at most n such that*

   $$\|f - P\|_\infty \leq \frac{c^k}{(n+1)n(n-1)\ldots(n-k+2)}\omega(f^{(k)}, 1/n).$$

Lemma 7. *For each $1 \le p < \infty$ there exists a constant $c_p > 0$ such that for any $d \ge 1$ there is a polynomial P of degree at most d such that $P(0) = 0$ and*

$$||t|^p - P(t)| \le \frac{c_p}{d^p} \quad \text{for all } t \in [-1, 1].$$

We have $c_p < (cp)^p$ for some universal $c > 0$.

Proof. By choosing c_p sufficiently large we may assume that $d > 2p$. Then $f_p(t) := |t|^p$ satisfies $f \in C^{(k)}[-1, 1]$ with $k = \lceil p \rceil - 1$, and

$$f^{(\lceil p \rceil - 1)}(t) = \text{sgn}^{\lceil p \rceil - 1}(t) p(p-1) \dots (p - \lceil p \rceil + 2)|t|^{p+1-\lceil p \rceil}.$$

Thus

$$\omega(f^{(\lceil p \rceil - 1)}, \delta) = p(p-1) \dots (p - \lceil p \rceil + 2)\delta^{p+1-\lceil p \rceil}.$$

From Jackson V we obtain a polynomial P of degree at most d with

$$
\begin{aligned}
\|f_p - P\|_\infty &\le \frac{c^{\lceil p \rceil - 1}}{(d+1)d(d-1)\dots(d - \lceil p \rceil + 3)} p(p-1)\dots(p - \lceil p \rceil + 2)(1/d)^{p+1-\lceil p \rceil} \\
&< \frac{c^{\lceil p \rceil - 1} p(p-1)\dots(p - \lceil p \rceil + 2)}{(\frac{1}{2}d)^{\lceil p \rceil - 1}}(1/d)^{p+1-\lceil p \rceil} \\
&= \frac{c}{d^p}.
\end{aligned}
$$

If we subtract the constant term from P we obtain a polynomial Q with $Q(0) = 0$ and $\|f_p - Q\|_\infty \le 2c/d^p$. $\qquad \square$

Smyth [33] approached Kusner's problem with the idea of approximating the p-norm with polynomials using Jackson's theorems. He obtained an upper bound $c_p n^{(p+1)/(p-1)}$ for $1 < p < \infty$ using linear independence. The next theorem was proved by Alon and Pudlák by combining Smyth's approach with the rank lemma.

Theorem 12 (Smyth [33], Alon & Pudlák [1]). *For $1 \le p < \infty$, $e(\ell_p^n) \le c_p' n^{(2p+2)/(2p-1)}$. We may take $c_p' = cp$, with $c > 0$ and absolute constant.*

Proof. Let $\boldsymbol{p}_1, \dots, \boldsymbol{p}_m \in \ell_p^n$ be 1-equilateral. Let $c = \max(c_p, (2^{1/p} - 1)^{-p})$, where c_p is the constant from Lemma 7. We fix an integer d such that $cn\sqrt{m} < d^p < 2cn\sqrt{m}$ (possible since $c \ge (2^{1/p} - 1)^{-p}$). Let P be the polynomial from Lemma 7. Define the $m \times m$ matrix $A = [a_{ij}]$ by $a_{ij} = 1 - \sum_{k=1}^n P(p_i^{(k)} - p_j^{(k)})$. Since $\sum_{k=1}^n P(p_i^{(k)} - p_j^{(k)})$ is an approximation of $\|\boldsymbol{p}_i - \boldsymbol{p}_j\|_p^p$, the matrix A is an approximation of the identity I_m. Again we estimate how close. Firstly, since $P(0) = 0$, all $a_{ii} = 1$. Secondly for any $i \ne j$,

$$
\begin{aligned}
|a_{ij}| &= \left| \|\boldsymbol{p}_i - \boldsymbol{p}_j\|_p^p - \sum_{k=1}^n P(p_i^{(k)} - p_j^{(k)}) \right| \\
&\le \sum_{k=1}^n \left| |\boldsymbol{p}_i^{(k)} - \boldsymbol{p}_j^{(k)}|^p - P(\boldsymbol{p}_i^{(k)} - \boldsymbol{p}_j^{(k)}) \right| \\
&\le n\frac{c_p}{d^p} < \frac{1}{\sqrt{m}},
\end{aligned}
$$

by choice of d. By Corollary 7 we have rank $A \ge m/2$.

We now find an upper bound for the rank of A. For each $i = 1, 2, \dots, m$, define the polynomial in n variables

$$P_i(x_1, \dots, x_n) := 1 - \sum_{k=1}^n P(p_i^{(k)} - x_i).$$

Thus $a_{ij} = P_i(\boldsymbol{p}_j)$. Each P_i is in the linear span of

$$\mathcal{P} = \{1, \sum_{k=1}^{n} x_i^d, x_1, \ldots, x_n, x_1^2, \ldots, x_n^2, \ldots, x_1^{d-1}, \ldots, x_n^{d-1}\}.$$

Thus the m polynomials P_i lie in a $(2 + (d-1)n)$-dimensional subspace of polynomials. Then the ith row vector of A, $\big[P_i(\boldsymbol{p}_1), \ldots, P_i(\boldsymbol{p}_m)\big]$, is in the linear span of

$$\{(f(\boldsymbol{p}_1), \ldots, f(\boldsymbol{p}_m)) : f \in \mathcal{P}\},$$

which is a subspace of $\mathbb{R}^m$ of dimension at most $|\mathcal{P}| = 2 + (d-1)n \leq dn$. It follows that $\operatorname{rank} A \leq dn$.

Putting the upper and lower bound for $\operatorname{rank} A$ together, we find $m/2 \leq dn$. Since $d^p < 2cn\sqrt{m}$, we obtain $m < c_p' n^{(2p+2)/(2p-1)}$. $\qquad\square$

In the next section we find the best known upper bounds for $e(\ell_1^n)$, due to Alon and Pudlák.

8 The best known upper bound for $e(\ell_1^n)$

Alon and Pudlák [1] proved that $e(\ell_p^n) \leq c_p n \log n$ if p is an odd integer. This matches the lower bound of $2n$ apart from the $\log n$ factor and the constant that depends on p. We here present their proof for the case $p = 1$. The proof for other odd p is simple once this case is understood, see [1] for the detail.

Theorem 13 (Alon and Pudlák [1]). $e(\ell_1^n) \leq cn \log n$.

Proof. Let $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{m+1}$ be 1-equilateral in ℓ_1^n. After a translation we may assume $\boldsymbol{p}_{m+1} = \boldsymbol{o}$. Thus it is sufficient to prove $m + 1 \leq cn \log n$ given that

$$\sum_{k=1}^{n} |p_i^{(k)}| \;=\; 1 \quad \forall i = 1, \ldots, m, \tag{17}$$

$$\text{and } \sum_{k=1}^{n} |p_i^{(k)}| \;=\; 1 \quad \forall 1 \leq i \neq j \leq m. \tag{18}$$

We first show that by doubling the dimension we may assume that all $p_i^{(k)} \geq 0$. We replace each $p_i^{(k)}$ by

$$\begin{cases} (p_i^{(k)}, 0) & \text{if } p_i^{(k)} \geq 0, \\ (0, -p_i^{(k)}) & \text{if } p_i^{(k)} \leq 0. \end{cases}$$

Now the points $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_m$ are in ℓ_1^{2n}, they are still 1-equilateral, $\|\boldsymbol{p}_i\|_1 = 1$, and now all $p_i^{(k)} \geq 0$. Thus we may assume together with (17) and (18) that $p_i^{(k)} \geq 0$ holds. (Once we have proven $m \leq cn \log n$ with this assumption, then $m \leq 1 + c(2n) \log(2n) < 3cn \log n$ in the general case.)

We now express $|p_i^{(k)} - p_j^{(k)}|$ in terms of $\min(p_i^{(k)}, p_j^{(k)})$. Since

$$|a - b| = a + b - 2\min(a, b) \quad \text{for } a, b \in \mathbb{R},$$

and $\sum_{k=1}^{n} p_i^{(k)} = 1$, we obtain

$$I_m = \big[1 - \|\boldsymbol{p}_i - \boldsymbol{p}_j\|\big]_{i,j=1}^{m} = \left[-1 + 2\sum_{k=1}^{n} \min\big(p_i^{(k)}, p_j^{(k)}\big)\right]_{i,j=1}^{m}.$$

We now want to approximate $\min(a, b)$ by an inner product $\langle \overline{\boldsymbol{a}}, \overline{\boldsymbol{b}} \rangle$ where $\overline{\boldsymbol{a}}, \overline{\boldsymbol{b}} \in \mathbb{R}^N$. Then I_m will be approximated by

$$A = \left[-1 + 2\left\langle \big(\overline{\boldsymbol{p}}_i^{(1)}, \ldots, \overline{\boldsymbol{p}}_i^{(n)}\big), \big(\overline{\boldsymbol{p}}_j^{(1)}, \ldots, \overline{\boldsymbol{p}}_j^{(n)}\big)\right\rangle\right]_{i,j=1}^{m}.$$

Then rank $A \leq 1 + Nn$, since it is a linear combination of J_m of rank 1 and a Gram matrix of vectors in $\mathbb{R}^{Nn}$. The approximation will be good enough so that the rank lemma will give rank $A > c_1 m$. Thus $c_1 m < Nn + 1$, and we will have to make sure that $N < c_2 \log n$. What will in fact happen is that N will be different for each coordinate $k = 1, \ldots, n$, say N_k, and we'll have to make sure $\sum_{k=1}^n N_k < c_2 n \log n$.

The rough idea for approximating $\min(a, b)$ by an inner product is the following. For each $x \in [0, 1]$ define

$$\overline{\boldsymbol{x}} = (\underbrace{\tfrac{1}{\sqrt{N}}, \ldots, \tfrac{1}{\sqrt{N}}}_{\lfloor xN \rfloor \text{ times}}, 0, \ldots, 0) \in \mathbb{R}^N.$$

Then for $0 \leq a \leq b \leq 1$,

$$\langle \overline{\boldsymbol{a}}, \overline{\boldsymbol{b}} \rangle = \tfrac{1}{N} \lfloor aN \rfloor = a - \varepsilon = \min(a, b) - \varepsilon,$$

where $0 \leq \varepsilon \leq 1/N$. Let us see how far this brings us already. Since

$$\left| \sum_{k=1}^n \min \left(p_i^{(k)}, p_j^{(k)} \right) - \left\langle \left(\overline{\boldsymbol{p}}_i^{(1)}, \ldots, \overline{\boldsymbol{p}}_i^{(n)} \right), \left(\overline{\boldsymbol{p}}_j^{(1)}, \ldots, \overline{\boldsymbol{p}}_j^{(n)} \right) \right\rangle \right| \leq \frac{n}{N},$$

the error by which A approximates I_m is

$$a_{ij} - \delta_{ij} \leq \frac{2n}{N} = \frac{1}{\sqrt{m}}$$

if we choose $N = 2n\sqrt{m}$. By the rank lemma, rank $A \geq (1 + o(1))m/2$. Also, rank $A \leq Nn + 1$. Putting the bounds together, we obtain $m < cn^4$, already obtained in Theorem 11.

We now fine-tune this idea. For any partition $\mathcal{P} = \{0 = u_0 < u_1 < \cdots < u_{N-1} < u_N = 1\}$ of $[0, 1]$ into N intervals, we define for any $x \in [0, 1]$ the vector $\boldsymbol{x}_{\mathcal{P}} \in \mathbb{R}^N$, where

$$x_{\mathcal{P}}^{(j)} = \begin{cases} \sqrt{u_j - u_{j-1}}, & j < t, \\ \frac{x - u_{t-1}}{\sqrt{u_t - u_{t-1}}}, & j = t, \\ 0, & j > t, \end{cases}$$

where t is such that $x \in (u_{t-1}, u_t)$.

Remarks.

1. We'll choose a different partition for each coordinate $k = 1, \ldots, n$. Since both $\sum_k \min(p_i^{(k)}, p_j^{(k)})$ and $\left\langle (\overline{\boldsymbol{p}}_i^{(1)}, \ldots, \overline{\boldsymbol{p}}_i^{(n)}), (\overline{\boldsymbol{p}}_j^{(1)}, \ldots, \overline{\boldsymbol{p}}_j^{(n)}) \right\rangle$ are sums over $k = 1, \ldots, n$, the error in each coordinate will also sum up, so we first do the analysis in a single coordinate.

2. We'll choose the partitions so that a coordinate will never hit an endpoint of any (u_{j-1}, u_j).

With the above definition we find for $0 \leq a \leq b \leq 1$ that

$$
\begin{aligned}
\langle \boldsymbol{a}_{\mathcal{P}}, \boldsymbol{b}_{\mathcal{P}} \rangle &= (\sqrt{u_1 - 0})^2 + (\sqrt{u_2 - u_1})^2 + (\sqrt{u_{t-1} - u_{t-2}})^2 + a_{\mathcal{P}}^{(t)} b_{\mathcal{P}}^{(t)}, \quad \text{where } a \in (u_{t-1}, u_t), \\
&= \begin{cases} u_{t-1} + \frac{a - u_{t-1}}{\sqrt{u_t - u_{t-1}}} \sqrt{u_t - u_{t-1}} & \text{if } b \notin (u_{t-1}, u_t), \\ u_{t-1} + \frac{a - u_{t-1}}{\sqrt{u_t - u_{t-1}}} \frac{b - u_{t-1}}{\sqrt{u_t - u_{t-1}}} & \text{if } b \in (u_{t-1}, u_t), \end{cases} \\
&= \begin{cases} \min(a, b) & \text{if } a \text{ and } b \text{ are in different intervals of } \mathcal{P}, \\ u_{t-1} + \frac{(a - u_{t-1})(b - u_{t-1})}{u_t - u_{t-1}} & \text{if } a, b \in (u_{t-1}, u_t). \end{cases}
\end{aligned}
$$

Thus there is no error if a and b are in different intervals. However, when a and b are in the same interval of $\mathcal{P}$, the error is bad, and we want to get rid of it. We do this by adding another N coordinates: Define

$$\widehat{\boldsymbol{x}}_{\mathcal{P}} = \frac{x - u_{t-1}}{\sqrt{u_t - u_{t-1}}} \boldsymbol{e}_t \in \mathbb{R}^N, \quad \text{with } t \text{ such that } x \in (u_{t-1}, u_t).$$

21

Then
$$\left\langle \widehat{\boldsymbol{a}}_{\mathcal{P}}, \widehat{\boldsymbol{b}}_{\mathcal{P}} \right\rangle = \begin{cases} 0 & \text{if } a \text{ and } b \text{ are in different intervals of } \mathcal{P}, \\ \frac{(a-u_{t-1})(b-u_{t-1})}{u_t - u_{t-1}} & \text{if } a, b \in (u_{t-1}, u_t). \end{cases}$$

Since we want to subtract this, we let the inner product on $\mathbb{R}^N \oplus \mathbb{R}^N$ be

$$\langle (\boldsymbol{x}_1, \boldsymbol{x}_2), (\boldsymbol{y}_1, \boldsymbol{y}_2) \rangle = \langle \boldsymbol{x}_1, \boldsymbol{y}_1 \rangle - \langle \boldsymbol{x}_2, \boldsymbol{y}_2 \rangle.$$

We then obtain for $0 \le a \le b \le 1$ that

$$\left\langle (\boldsymbol{a}_{\mathcal{P}}, \widehat{\boldsymbol{a}}_{\mathcal{P}}), (\boldsymbol{b}_{\mathcal{P}}, \widehat{\boldsymbol{b}}_{\mathcal{P}}) \right\rangle = \begin{cases} \min(a,b) & \text{if } a \text{ and } b \text{ are in different intervals of } \mathcal{P}, \\ u_{t-1} & \text{if } a, b \in (u_{t-1}, u_t). \end{cases}$$

Thus if a and b are in the same interval, the approximation is a systematic rounding down. We now add a third vector in $\mathbb{R}^N$ to improve the approximation in this case. We do this by *randomized rounding*: For each interval (u_{t-1}, u_t) we choose randomly and uniformly $\tau \in (u_{t-1}, u_t)$ (a *threshold*), and for any $x \in (u_{t-1}, u_t)$ we define

$$\widetilde{\boldsymbol{x}}_{\mathcal{P}} = \begin{cases} 0 & \text{if } x \le \tau, \\ \sqrt{u_t - u_{t-1}}\, \boldsymbol{e}_t & \text{if } x > \tau. \end{cases}$$

Note that $\widetilde{\boldsymbol{x}}_{\mathcal{P}}$ is a random variable. Thus if a and b are in different intervals then $\left\langle \widetilde{\boldsymbol{a}}_{\mathcal{P}}, \widetilde{\boldsymbol{b}}_{\mathcal{P}} \right\rangle = 0$, and if a and b are in the same interval (u_{t-1}, u_t), then (as with $\widehat{\boldsymbol{x}}_{\mathcal{P}}$)

$$\left\langle \widetilde{\boldsymbol{a}}_{\mathcal{P}}, \widetilde{\boldsymbol{b}}_{\mathcal{P}} \right\rangle = \begin{cases} 0 & \text{if } \min(a,b) \le \tau, \\ u_t - u_{t-1} & \text{if } \min(a,b) > \tau. \end{cases}$$

We now let the inner product on $\mathbb{R}^N \oplus \mathbb{R}^N \oplus \mathbb{R}^N$ be

$$\langle (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3), (\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3) \rangle = \langle \boldsymbol{x}_1, \boldsymbol{y}_1 \rangle - \langle \boldsymbol{x}_2, \boldsymbol{y}_2 \rangle + \langle \boldsymbol{x}_3, \boldsymbol{y}_3 \rangle,$$

and for each $x \in [0, 1]$, let
$$\overline{\boldsymbol{x}}_{\mathcal{P}} = (\boldsymbol{x}_{\mathcal{P}}, \widehat{\boldsymbol{x}}_{\mathcal{P}}, \widetilde{\boldsymbol{x}}_{\mathcal{P}}) \in \mathbb{R}^{3N}.$$

Then for any $a, b \in [0, 1]$,

$$\langle \overline{\boldsymbol{a}}_{\mathcal{P}}, \overline{\boldsymbol{b}}_{\mathcal{P}} \rangle = \begin{cases} \min(a,b) & \text{if } a, b \text{ are in different intervals of } \mathcal{P}, \\ \begin{cases} u_{t-1}, & \text{if } \min(a,b) \le \tau \\ u_t, & \text{if } \min(a,b) > \tau \end{cases}, & \text{if } a, b \in (u_{t-1}, u_t), \end{cases}$$

and the error $X = \min(a,b) - \langle \overline{\boldsymbol{a}}_{\mathcal{P}}, \overline{\boldsymbol{b}}_{\mathcal{P}} \rangle$ (depending on $\mathcal{P}, a, b$) satisfies

$$X = \begin{cases} 0 & \text{if } a, b \text{ in different intervals,} \\ a - u_{t-1} & \text{if } \min(a,b) \le \tau, \ a, b \in (u_{t-1}, u_t), \\ a - u_t & \text{if } \min(a,b) > \tau, \ a, b \in (u_{t-1}, u_t). \end{cases}$$

Thus
$$|X| \le u_t - u_{t-1} = \text{length}(u_{t-1}, u_t),$$

the length of the interval. We now calculate the expected value and variance of X. Without loss of generality $a \le b$. Then easily

$$\begin{aligned} E(X) &= \text{Prob}(a \le \tau)(a - u_{t-1}) + \text{Prob}(a > \tau)(a - u_{t-1}) \\ &= \frac{u_t - a}{u_t - u_{t-1}}(a - u_{t-1}) + \frac{a - u_{t-1}}{u_t - u_{t-1}}(a - u_t) \\ &= 0, \end{aligned}$$

and

$$
\begin{aligned}
E(X^2) &= \operatorname{Prob}(a \le \tau)(a - u_{t-1})^2 + \operatorname{Prob}(a > \tau)(a - u_{t-1})^2 \\
&= \frac{u_t - a}{u_t - u_{t-1}}(a - u_{t-1})^2 + \frac{a - u_{t-1}}{u_t - u_{t-1}}(a - u_t)^2 \\
&= \frac{(a - u_{t-1})(u_t - a)}{u_t - u_{t-1}}(a - u_{t-1} - a + u_t) \\
&= (a - u_{t-1})(u_t - a) \\
&\le \frac{1}{4}(u_t - u_{t-1})^2, \text{ since } u_{t-1} < a < u_t.
\end{aligned}
$$

For each coordinate $k = 1, \dots, n$, we'll (soon) choose a different partition $\mathcal{P}_k$ of $[0, 1]$ into N_k intervals, with an independent random threshold for each interval in each $\mathcal{P}_k$. For each $\boldsymbol{x} = (x(1), \dots, x(n))$ we let

$$
\overline{\boldsymbol{x}} = (\overline{\boldsymbol{x}(1)}_{\mathcal{P}_1}, \dots, \overline{\boldsymbol{x}(n)}_{\mathcal{P}_n}) \in \mathbb{R}^{3N_1} \oplus \dots \oplus \mathbb{R}^{3N_n}.
$$

Thus for each $\boldsymbol{p}_i$ we now have a $\overline{\boldsymbol{p}}_i = (\overline{\boldsymbol{p}}_i^{(1)}, \dots, \overline{\boldsymbol{p}}_i^{(n)}) \in \mathbb{R}^{3 \sum_{k=1}^n N_k}$ (where in each coordinate a different partition is used, but not denoted anymore). Then we approximate the identity

$$
I_m = \left[-1 + 2 \sum_{k=1}^n \min\left(p_i^{(k)}, p_j^{(k)} \right) \right]_{i,j=1}^m
$$

by

$$
A = \left[-1 + 2 \left\langle \overline{\boldsymbol{p}}_i, \overline{\boldsymbol{p}}_j \right\rangle \right]_{i,j=1}^m = \left[-1 + 2 \sum_{k=1}^n \left\langle \overline{\boldsymbol{p}}_i^{(k)}, \overline{\boldsymbol{p}}_j^{(k)} \right\rangle \right]_{i,j=1}^m
$$

and we let $X_{i,j,k} = \min\left(p_i^{(k)}, p_j^{(k)} \right) - \left\langle \overline{\boldsymbol{p}}_i^{(k)}, \overline{\boldsymbol{p}}_j^{(k)} \right\rangle$. Note that $\{ X_{i,j,k} : k = 1, \dots, n \}$ are independent random variables for fixed i, j, since $X_{i,j,k} \equiv 0$ if $p_i^{(k)}$ and $p_j^{(k)}$ are in different intervals of $\mathcal{P}_k$, or depends on a single threshold, with these thresholds independent as $k = 1, \dots, n$.

We want to apply the rank lemma to A, so we start estimating $\left(\sum_i a_{ii} \right)^2$ and $\sum_{i,j} a_{ij}^2$. First of all,

$$
\begin{aligned}
a_{ii} &= -1 + 2 \sum_{k=1}^n \left\langle \overline{\boldsymbol{p}}_i^{(k)}, \overline{\boldsymbol{p}}_i^{(k)} \right\rangle \\
&= -1 + 2 \sum_{k=1}^n \min\left(p_i^{(k)}, p_i^{(k)} \right) - 2 \sum_{k=1}^n X_{i,i,k} \\
&= 1 - 2 \sum_{k=1}^n X_{i,i,k}.
\end{aligned}
$$

Similarly, for $i \ne j$,

$$
a_{ij} = -2 \sum_{k=1}^n X_{i,j,k}.
$$

Since $|X_{i,i,k}| \le \operatorname{length}(I_k)$, where I_k is the interval of $\mathcal{P}_k$ containing $p_i^{(k)}$, we obtain

$$
\sum_{k=1}^n X_{i,i,k} \le \sum_{k=1}^n \operatorname{length}(I_k),
$$

and

$$\sum_{i=1}^{m} a_{ii} \;=\; m - 2\sum_{i=1}^{m}\sum_{k=1}^{n} X_{k,i,i}$$

$$\geq\; m - 2\sum_{i=1}^{m}\sum_{k=1}^{n} \mathrm{length}(I_k)$$

$$=\; m - 2\sum_{I} \mathrm{hit}(I)\,\mathrm{length}(I), \qquad (19)$$

wehre the last sum is over all intervals of all $\mathcal{P}_k$, and $\mathrm{hit}(I)$ is the number of $p_i^{(k)}$ "hitting" I, i.e., the number of ordered pairs (i,k) with $p_i^{(k)} \in I$.

Secondly, for $i \neq j$, $a_{ij}^2 = 4(\sum_{k=1}^{n} X_{i,j,k})^2$, hence $E(a_{ij}^2) = 4\sum_{k=1}^{n} E(X_{i,j,k}^2)$ due to independence and $E(X_{i,j,k}) = 0$. Similarly,

$$E(a_{ii}^2) \;=\; 1 - 4\sum_{k=1}^{n} E(X_{i,i,k}) + 4\sum_{k=1}^{n} E(X_{i,i,k}^2)$$

$$=\; 1 + 4\sum_{k=1}^{n} E(X_{i,i,k}^2).$$

Thus

$$E\Big(\sum_{i,j=1}^{m} a_{ij}^2\Big) \;=\; \sum_{i=1}^{m} E(a_{ii}^2) + \sum_{\substack{i,j=1 \\ i\neq j}}^{m} E(a_{ij}^2)$$

$$=\; m + 4\sum_{i=1}^{m}\sum_{k=1}^{n} E(X_{i,i,k}^2) + 4\sum_{\substack{i,j=1 \\ i\neq j}}^{m}\sum_{k=1}^{n} E(X_{i,j,k}^2)$$

$$=\; m + 4\sum_{i,j=1}^{m}\sum_{k=1}^{n} E(X_{i,j,k}^2)$$

$$\leq\; m + \sum_{i,j=1}^{m}\sum_{k=1}^{n} \mathrm{length}(I_{i,j,k})^2, \quad \text{with } I_{i,j,k} \in \mathcal{P}_k \text{ such that } p_i^{(k)}, p_j^{(k)} \in I_{i,j,k} \text{ if} $$
$$\text{it exists, otherwise take } \mathrm{length}(I_{i,j,k}) = 0,$$

$$=\; m + \sum_{I} \mathrm{hit}(I)^2\,\mathrm{length}(I)^2,$$

where the last sum is again over all intervals. Because we have found an upper bound on the expected value of $\sum_{i,j=1}^{m} a_{ij}^2$, it follows that there exists a choice of thresholds such that

$$\sum_{i,j=1}^{m} a_{ij}^2 \leq m + \sum_{I} \mathrm{hit}(I)^2\,\mathrm{length}(I)^2. \qquad (20)$$

Fix such a choice of thresholds. (So at this stage we leave randomness behind.) The rank lemma together with (19) and (20) gives

$$\mathrm{rank}\,A \geq \frac{(m - 2\sum_I \mathrm{hit}(I)\,\mathrm{length}(I))^2}{m + \sum_I \mathrm{hit}(I)^2\,\mathrm{length}(I)^2}$$

if

$$m - 2\sum_{I} \mathrm{hit}(I)\,\mathrm{length}(I) \geq 0. \qquad (21)$$

24

On the other hand, rank $A \leq 1 + \sum_{k=1}^{n} 3N_k = 1 + 3\sum_I 1$, thus

$$\frac{(m - 2\sum_I \text{hit}(I)\,\text{length}(I))^2}{m + \sum_I \text{hit}(I)^2\,\text{length}(I)^2} \leq 1 + 3\sum_I 1. \tag{22}$$

Finally, we fix the partition $\mathcal{P}_k$ for each coordinate k. We assume that for each $k = 1, \ldots, n$, all $p_i^{(k)}$ are distinct. (This is not essential: the $p_i^{(k)}$ may all be perturbed by a sufficiently small amount so as to weaken the left-hand side of (22) by at most $\varepsilon > 0$, for any $\varepsilon > 0$.)

We now assume that both m and n are powers of 4. If we can prove $m \leq cn\log n$ in this case, then $m \leq 16cn\log n$ in general, since we may round m down to a power of 4 (thus dividing m by at most 4), and n up to a power of 4 (thus multiplying n by at most 4). All the logarithms until the end of the proof will be base 2 (and indicated as such). We first divide $[0, \frac{1}{n}]$ into $\sqrt{m/n}$ equal parts (thus each of length $1/\sqrt{mn}$). These are called *base intervals*. Without loss of generality no $p_i^{(k)}$ is an endpoint of a base interval, again by making an infinitesimal perturbation.

Then start at $t = 1$ and let t go down until

$$\text{hit}([t, 1])\,\text{length}([t, 1]) \geq \frac{cn\lg n}{m}$$

or until $t = 2^{-1/3}$, whichever comes first. (The constant c will be fixed later; $c = 10000$ is sufficient.) If the stopping point t_1 equals some $p_i^{(k)}$, we go down slightly more, without increasing $\text{hit}([t, 1])$. Then start with a new interval at t_1 and go down until

$$\text{hit}([t, t_1))\,\text{length}([t, t_1]) \geq \frac{cn\lg n}{m}$$

or until $t = 2^{-2/3}$, whichever comes first. In general we go down until

$$\text{hit}([t, t_{s-1}))\,\text{length}([t, t_{s-1})) \geq \frac{cn\lg n}{m}$$

or until $t = 2^{-s/3}$, for each $s = 1, \ldots, 3\lg n$.

If $\text{hit}(I)\,\text{length}(I) \geq cn\lg n/m$, we call I a *regular interval*, otherwise a *singular interval*.

Thus for a singular interval I we have

$$\text{hit}(I)\,\text{length}(I) < \frac{cn\lg n}{m}, \quad I \text{ singular.}$$

For a regular interval, if the stopping point did not hit a $p_i^{(k)}$, then $\text{hit}(I)\,\text{length}(I) = cn\lg n/m$. If the stopping point did hit a $p_i^{(k)}$, then $\text{hit}(I)\,\text{length}(I) > cn\lg n/m$, but

$$(\text{hit}(I) - 1)\,\text{length}(I) \leq \frac{cn\lg n}{m}.$$

Thus

$$\text{hit}(I)\,\text{length}(I) \geq \frac{cn\lg n}{m}, \quad I \text{ regular,}$$

and

$$\text{hit}(I)\,\text{length}(I) \leq \frac{2cn\lg n}{m} \quad \text{if } \text{hit}(I) \geq 2 \text{ for regular } I.$$

Later we'll have to take care of regular I with $\text{hit}(I) = 1$ separately: for them

$$\sum_{\substack{I \text{ regular} \\ \text{hit}(I)=1}} \text{length}(I)^2 \leq \sum_{\substack{I \text{ regular} \\ \text{hit}(I)=1}} \text{length}(I) < \sum_I \text{length}(I) = n.$$

25

There are at most $3n \lg n$ singular intervals. We now bound the number of regular intervals. For each regular interval I, let s_I be the s for which $I \subseteq [2^{-s/3}, 2^{-(s-1)/3}]$. Then

$$
\begin{aligned}
\sum_{\text{regular } I} 2^{-s_I/3} \operatorname{hit}(I) \;\;&\leq\;\; \sum_{\text{regular } I} \min(I) \operatorname{hit}(I) \\[2mm]
&<\;\; \sum_{I} \sum_{\substack{i=1 \\ p_i^{(k)} \in I}}^{m} \min(I), \quad \text{where } k \text{ is such that } I \in \mathcal{P}_k, \\[2mm]
&<\;\; \sum_{I} \sum_{\substack{i=1 \\ p_i^{(k)} \in I}}^{m} p_i^{(k)} = \sum_{i=1}^{m} \sum_{k=1}^{n} p_i^{(k)} = m.
\end{aligned}
\tag{23}
$$

We now bound $\sum_{\text{regular } I} 1/(2^{-s_I/3} \operatorname{hit}(I))$. For each $s = 1, \ldots, 3 \lg n$,

$$
\begin{aligned}
\sum_{\substack{\text{regular } I \\ s_I = s}} \operatorname{length}(I) \;\;&\leq\;\; (2^{-(s-1)/3} - 2^{-s/3})n \quad \text{(since for each coordinate the } I\text{'s don't overlap)} \\[2mm]
&=\;\; (2^{1/3} - 1)2^{2/3}n < \tfrac{1}{3}2^{-s/3}n.
\end{aligned}
$$

Since $\operatorname{length}(I) \operatorname{hit}(I) \geq cn \lg n / m$ for regular I,

$$
\frac{1}{\operatorname{hit}(I)} \leq \frac{m \operatorname{length}(I)}{cn \lg n}.
$$

Thus

$$
\begin{aligned}
\sum_{\substack{\text{regular } I \\ s_I = s}} \frac{1}{2^{-s/3} \operatorname{hit}(I)} \;\;&\leq\;\; \frac{m}{2^{-s/3} cn \lg n} \sum_{\substack{\text{regular } I \\ s_I = s}} \operatorname{length}(I) \\[2mm]
&<\;\; \frac{m}{3c \lg n}.
\end{aligned}
$$

Thus

$$
\sum_{\text{regular } I} \frac{1}{2^{-s_I/3} \operatorname{hit}(I)} < \sum_{s=1}^{3 \lg n} \frac{m}{3c \lg n} = \frac{m}{c}.
\tag{24}
$$

By Cauchy-Schwartz,

$$
\begin{aligned}
\left(\sum_{\text{regular } I} 1 \right)^2 \;\;&=\;\; \left(\sum_{\text{regular } I} \sqrt{2^{-s_I/3} \operatorname{hit}(I)} \frac{1}{\sqrt{2^{-s_I/3} \operatorname{hit}(I)}} \right)^2 \\[2mm]
&\leq\;\; \left(\sum_{\text{regular } I} 2^{-s_I/3} \operatorname{hit}(I) \right) \left(\sum_{\text{regular } I} \frac{1}{2^{-s_I/3} \operatorname{hit}(I)} \right) \\[2mm]
&<\;\; m\frac{m}{c} \quad \text{by (23) and (24)},
\end{aligned}
$$

and the number of regular intervals is less than $m/\sqrt{c}$. Thus the total number of non-base intervals is at most $3n \lg n + m/\sqrt{c}$ (and recall that the number of base intervals is $\sqrt{mn}$). We are now in a position to estimate the various quantities in (22). We want to show that these estimates imply $m \leq cn \lg n$, so we

assume that $m > cn \lg n$ and aim for a contradiction if c is sufficiently large (10000 will do). First of all,

$$
\begin{aligned}
1 + 3 \sum_I 1 \quad &< \quad 1 + 3 \left(\sqrt{mn} + 3n \lg n + \frac{m}{\sqrt{c}} \right) \\
&< \quad 1 + \frac{3m}{\sqrt{c \lg n}} + \frac{9m}{c} + \frac{3m}{\sqrt{c}} \\
&< \quad \frac{16m}{\sqrt{c}}
\end{aligned}
\tag{25}
$$

to be sure. Secondly,

$$
\begin{aligned}
m + \sum_I \operatorname{hit}(I)^2 \operatorname{length}(I)^2 \quad = \quad & m + \sum_{\substack{\text{base } I}} \operatorname{hit}(I)^2 \frac{1}{mn} + \sum_{\substack{\text{regular } I \\ \operatorname{hit}(I)=1}} \operatorname{length}(I)^2 \\
& + \sum_{\substack{\text{non-base } I \\ h_I \geq 2 \text{ if regular}}} (\operatorname{hit}(I) \operatorname{length}(I))^2 \\
< \quad & m + nm^2 \frac{1}{mn} + n + \left(3n \lg n + \frac{m}{\sqrt{c}} \right) \left(\frac{2cn \lg n}{m} \right)^2 \\
< \quad & 2m + \frac{m}{c \lg n} + \frac{12m}{c} + \frac{4m}{\sqrt{c}} \\
< \quad & 2m + \frac{17m}{\sqrt{c}}.
\end{aligned}
\tag{26}
$$

Thirdly,

$$
\begin{aligned}
\sum_I \operatorname{hit}(I) \operatorname{length}(I) \quad = \quad & \sum_{\substack{\text{base } I}} \operatorname{hit}(I) \frac{1}{\sqrt{mn}} + \sum_{\substack{\text{regular } I \\ \operatorname{hit}(I)=1}} \operatorname{length}(I) \\
& + \sum_{\substack{\text{non-base } I \\ h_I \geq 2 \text{ if regular}}} \operatorname{hit}(I) \operatorname{length}(I) \\
< \quad & nm \frac{1}{\sqrt{mn}} + n + \left(3n \lg n + \frac{m}{\sqrt{c}} \right) \left(\frac{2cn \lg n}{m} \right) \\
< \quad & \frac{m}{\sqrt{c \lg n}} + \frac{m}{c \lg n} + \frac{6m}{c} + \frac{2m}{\sqrt{c}} \\
< \quad & \frac{10m}{\sqrt{c}}.
\end{aligned}
$$

Thus if $\sqrt{c} \geq 20$, then (21) is satisfied, and in fact

$$
m - 2 \sum_I \operatorname{hit}(I) \operatorname{length}(I) > m - \frac{10m}{\sqrt{c}}.
\tag{27}
$$

Substituting estimates (25), (26), and (27) into (22) we obtain

$$
\frac{16m}{\sqrt{c}} > \frac{\left(m - \frac{10m}{\sqrt{c}} \right)^2}{2m + \frac{17m}{\sqrt{c}}}, \quad \text{i.e.,} \quad \frac{16}{\sqrt{c}} > \frac{\left(1 - \frac{10}{\sqrt{c}} \right)^2}{2 + \frac{17}{\sqrt{c}}},
$$

which is false if c is sufficiently large, e.g. for $\sqrt{c} = 100$.

Thus $m < cn \lg n$, and the proof is finished. $\qquad \square$

27

9 Final remarks

9.1 Infinite dimensions

In infinite dimensions almost nothing is known. The most important open question here is the following.

Problem 6. *Does there exist an infinite equilateral set in any separable infinite-dimensional normed space?*

As observed by Terenzi [38], it follows from the partition calculus of set theory that a normed space of cardinality at least 2^c has an infinite equilateral set (in fact an uncountable one), where $c = 2^{\aleph_0}$ is the cardinality of the continuum. The following two easy results are proved in [35].

Theorem 14. *Let $\varepsilon > 0$ and X an infinite dimensional Banach space. Then X has an equivalent norm with Banach-Mazur distance of at most $2(1+\varepsilon)$ to the original norm, admitting an infinte equilateral set.*

Theorem 15. *Let $\varepsilon > 0$ and X an infinite dimensional superreflexive Banach space. Then X has an equivalent norm with Banach-Mazur distance of at most $1 + \epsilon$ to the original norm, admitting an infinite equilateral set.*

9.2 Generalizations

An *ε-almost-equilateral set* in a normed space is a set S satisfying $1 - \varepsilon \leq \|\boldsymbol{x} - \boldsymbol{y}\| \leq 1 + \varepsilon$ for all $\boldsymbol{x}, \boldsymbol{y} \in S$, $\boldsymbol{x} \neq \boldsymbol{y}$. Such sets have been studied in [22, 9, 7, 6, 5, 2].

Problem 7. *Prove that for each $\varepsilon > 0$ there exists $\delta > 0$ such that each n-dimensional normed space has an ε-almost-equilateral set of size $(1 + \delta)^n$.*

Equilateral sets may be generalized to the notion of a k-distance set, i.e., a subset S of a metric space such that $\{d(x, y) : x, y \in S, x \neq y\}$ consists of at most k numbers. Then the Euclidean case is already non-trivial, and for normed spaces a few results are known: see [36]. The following conjecture would generalize Petty's theorem:

Conjecture 3 ([36]). *A k-distance set in an n-dimensional normed space has size at most $(k+1)^n$. If equality holds for some $k \geq 1$, then the space must be isometric to ℓ_∞^n.*

It is known to be true for $n = 2$ and for ℓ_∞^n, while for general Minkowski spaces there are some weaker estimates; see [36].

Acknowledgements

References

[1] N. Alon and P. Pudlák, *Equilateral sets in l_p^n*, Geom. Funct. Anal. **13** (2003), no. 3, 467–482. MR 1 995 795

[2] Juan Arias-de Reyna, Keith Ball, and Rafael Villa, *Concentration of the distance in finite-dimensional normed spaces*, Mathematika **45** (1998), no. 2, 245–252. MR 2000b:46013

[3] H.-J. Bandelt, V. Chepoi and M. Laurent, *Embedding into rectilinear spaces*, Discrete Comput. Geom **19** (1998), 595–604. MR 99d:51017

[4] I. Bárány, *The densest $(n + 2)$-set in R^n*, Intuitive Geometry (1991), Coll. Math. Soc. János Bolyai 63, Elsevier, Amsterdam, 1994, pp. 7–10. MR 97a:52012

[5] J. Bastero, A. Peña, and G. Schechtman, *Embedding l_∞^n-cubes in low-dimensional Schatten classes*, Geometric aspects of functional analysis (Israel, 1992–1994), Oper. Theory Adv. Appl., vol. 77, Birkhäuser, Basel, 1995, pp. 5–11. MR 96j:46006

[6] Jesús Bastero and Julio Bernués, *Applications of deviation inequalities on finite metric sets*, Math. Nachr. **153** (1991), 33–41. MR 93a:46017

[7] Jesús Bastero, Julio Bernués, and Nigel Kalton, *Embedding l_∞^n-cubes in finite-dimensional 1-subsymmetric spaces*, Rev. Mat. Univ. Complut. Madrid **2** (1989), no. suppl., 47–52, Congress on Functional Analysis (Madrid, 1988). MR 91h:46024

[8] Leonard M. Blumenthal, *Theory and applications of distance geometry*, Second edition, Chelsea Publishing Co., New York, 1970. MR 42 #3678

[9] J. Bourgain, V. Milman, and H. Wolfson, *On type of metric spaces*, Trans. Amer. Math. Soc. **294** (1986), no. 1, 295–317. MR 88h:46033

[10] Peter Braß, *On equilateral simplices in normed spaces*, Beiträge Algebra Geom. **40** (1999), no. 2, 303–307. MR 2000i:52012

[11] E. W. Cheney, *Introduction to approximation theory*, AMS Chelsea Publishing, Providence, RI, 1998, Reprint of the second (1982) edition. MR 99f:41001

[12] L. Danzer and B. Grünbaum, *Über zwei Probleme bezüglich konvexer Körper von P. Erdős und von V. L. Klee*, Math. Z. **79** (1962), 95–99. MR 25 #1488

[13] B. V. Dekster, *Simplexes with prescribed edge lengths in Minkowski and Banach spaces*, Acta Math. Hungar. **86** (2000), no. 4, 343–358. MR 2001b:52001

[14] B. V. Dekster and J. B. Wilker, *Edge lengths guaranteed to form a simplex*, Arch. Math. (Basel) **49** (1987), no. 4, 351–366. MR 88k:51035

[15] Z. Füredi, J. C. Lagarias, and F. Morgan, *Singularities of minimal surfaces and networks and related extremal problems in Minkowski space*, Discrete and computational geometry (New Brunswick, NJ, 1989/1990), DIMACS Ser. Discrete Math. Theoret. Comput. Sci., vol. 6, Amer. Math. Soc., Providence, RI, 1991, pp. 95–109. MR 93d:52009

[16] Yehoram Gordon, *Some inequalities for Gaussian processes and applications*, Israel J. Math. **50** (1985), no. 4, 265–289. MR 87f:60058

[17] Helmut Groemer, *Abschätzungen für die Anzahl der konvexen Körper, die einen konvexen Körper berühren*, Monatsh. Math. **65** (1961), 74–81. MR 23 #A2129

[18] Branko Grünbaum, *On a conjecture of H. Hadwiger*, Pacific J. Math. **11** (1961), 215–219. MR 25 #1492

[19] Branko Grünbaum, *Strictly antipodal sets*, Israel J. Math. **1** 1963 5–10. MR 28 #2480

[20] R. K. Guy, *An olla-podrida of open problems, often oddly posed*, Amer. Math. Monthly **90** (1983), 196–199.

[21] Tao Jiang and Alexander Vardy, *Asymptotic Improvement of the Gilbert-Varshamov Bound on the Size of Binary Codes*, arXiv:math.CO/0404325, to appear in IEEE Transactions on Information Theory, 2004.

[22] William B. Johnson and Joram Lindenstrauss, *Extensions of Lipschitz mappings into a Hilbert space*, Conference in modern analysis and probability (New Haven, Conn., 1982), Contemp. Math., vol. 26, Amer. Math. Soc., Providence, RI, 1984, pp. 189–206. MR 86a:46018

[23] Jack Koolen, Monique Laurent and Alexander Schrijver, *Equilateral dimension of the rectilinear space*, Designs, Codes Cryptogr. **21** (2000), 149–164. MR 2001j:52013

[24] Gary Lawlor and Frank Morgan, *Paired calibrations applied to soap films, immiscible fluids, and surfaces or networks minimizing other norms*, Pacific J. Math. **166** (1994), no. 1, 55–83. MR 95i:58051

[25] J. H. van Lint, *Introduction to Coding Theory*, third ed., Springer-Verlag, Berlin, 1992.

[26] Horst Martini, Konrad J. Swanepoel, and Gunter Weiß, *The geometry of Minkowski spaces—a survey. I*, Expo. Math. **19** (2001), no. 2, 97–142. MR 2002h:46015a. Erratum Expo. Math. **19** (2001), no. 4, 364. MR 2002h:46015b

[27] K. Menger, *Untersuchungen über allgemeine Metrik*, Math. Ann. **100** (1928), 75–163.

[28] Frank Morgan, *Minimal surfaces, crystals, shortest networks, and undergraduate research*, Math. Intelligencer **14** (1992), no. 3, 37–44. MR 93h:53012

[29] C. M. Petty, *Equilateral sets in Minkowski spaces*, Proc. Amer. Math. Soc. **29** (1971), 369–374. MR 43 #1051

[30] K. Schütte, *Minimale Durchmesser endlicher Punktmengen mit vorgeschriebenem Mindestabstand*, Math. Ann. **150** (1963), 91–98. MR 26 #5479

[31] J. J. Seidel, *Quasiregular two-distance sets*, Indag. Math **31** (1969), 64–69.

[32] C. E. Shannon, *Probability of error for optimal codes in a Gaussian channel*, Bell System Tech. J. **38** (1959), 611–656.

[33] Cliff Smyth, *Equilateral or 1-distance sets and Kusner's conjecture*, manuscript, 2002.

[34] P. S. Soltan, *Analogues of regular simplexes in normed spaces*, Dokl. Akad. Nauk SSSR **222** (1975), no. 6, 1303–1305, English translation: Soviet Math. Dokl. 16 (1975), no. 3, 787–789. MR 52 #4127

[35] K. J. Swanepoel, *Combinatorial geometry of Minkowski spaces*, PhD Thesis, University of Pretoria, 1997.

[36] K. J. Swanepoel, *Cardinalities of k-distance sets in Minkowski spaces*, Discrete Math. **197/198** (1999), 759–767, 16th British Combinatorial Conference (London, 1997). MR 99k:52028

[37] K. J. Swanepoel, *A problem of Kusner on equilateral sets*, to appear in Archiv der Mathematik, 2004.

[38] Paolo Terenzi, *Equilater sets in Banach spaces*, Boll. Un. Mat. Ital. A (7) **3** (1989), no. 1, 119–124. MR 90c:46017

[39] A. C. Thompson, *Minkowski geometry*, Encyclopedia of Mathematics and its Applications, vol. 63, Cambridge University Press, Cambridge, 1996. MR 97f:52001

[40] A. D. Wyner, *Capabilities of bounded discrepancy decoding*, Bell System Tech. J. **44** (1965), 1061–1122.